

586

Db 271 GWCAINSKILVPPRTTHAKFTDWKYLKRLVSGRTLPVDFHIKXVSMKIPFQGMRL 330  
Qy 121 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGK 180  
Db 331 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGK 390  
Qy 181 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGGMFECKMKEADPLNLGNICVATV 240  
Db 391 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGGMFECKMKEADPLNLGNICVATV 450  
Qy 241 CKVLLDGYLMICVDGGPSTGLDFWCHASSHAIFPATFCQKNDIELTPPKGYEATFNW 300  
Db 451 CKVLLDGYLMICVDGGPSTGLDFWCHASSHAIFPATFCQKNDIELTPPKGYEATFNW 510  
Qy 301 ENYLEKTSKAAPSRLEFNMDCPNHGFVKGMKLEAVDLMEPRLICVATVKKVHRLLSIHF 360  
Db 511 ENYLEKTSKAAPSRLEFNMDCPNHGFVKGMKLEAVDLMEPRLICVATVKKVHRLLSIHF 570  
Qy 361 DGWSEYDQWVDCESPDIPYVGCWELTGYQLQPPVAAPATPLKAKEATKKKKQFGKKR 420  
Db 571 DGWSEYDQWVDCESPDIPYVGCWELTGYQLQPPVAAPATPLKAKEATKKKKQFGKKR 630  
Qy 421 KRIPPTKTRPLRQSGKXPLEDDPQGARKISSEVPVGEIIAVRVKESHLDVASPDKASSP 480  
Db 631 KRIPPTKTRPLRQSGKXPLEDDPQGARKISSEVPVGEIIAVRVKESHLDVASPDKASSP 690  
Qy 481 ELPVSVENIKQETDD 495  
Db 691 ELPVSVENIKQETDD 705

RESULT 2  
LML2 MOUSE  
ID LML2 MOUSE STANDARD; PRT; 703 AA.  
EC PS9178;  
ET 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Lethal(3)malignant brain tumor-like 2 protein (L(3)mblt-like 2 protein)  
DE (H-1(3)mblt-like protein).  
GN L3MBTL2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Liver;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences."  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RL -1- FUNCTION: Putative Polycarb group (pcg) protein. PcG proteins  
CC maintain the transcriptionally repressive state of genes, probably  
CC via a modification of chromatin, rendering it heritably changed in

CC its expressibility. Its association with a chromatin remodeling  
CC complex suggests that it may contribute to prevent expression of  
CC genes that trigger the cell into mitosis (by similarity).  
CC -1- SUBUNIT: Part of the E2F6.com-1 complex in G0 phase composed of  
CC E2F6, MGA, MAX, TFDPl, CBX3, BAT8, E2F6, RING1, RNF2, MELR,  
CC BAT8 and YAF2 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -1- SIMILARITY: Contains 1 C4-type zinc finger.  
CC -1- SIMILARITY: Contains 4 MBT domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; BC030864; AAH30864.1;  
DR MGD; MGI:2443584; 4732493N06Rik.  
DR SMART; SM00561; MBT; 4.  
KW Transcription regulation; Chromatin regulator; Nuclear protein;  
KW Zinc-finger; Metal-binding; Repeat.  
FT ZN FING 90 114 C4-TYPE.  
FT DOMAIN 214 291 MBT 1.  
FT DOMAIN 327 399 MBT 2.  
FT DOMAIN 432 509 MBT 3.  
FT DOMAIN 540 611 MBT 4.  
FT DOMAIN 17 21 POLY-GLU.  
FT DOMAIN 620 624 POLY-LYS.  
SQ SEQUENCE 703 AA; 78954 MW; 259B5DF5BFAAE9 CRC64;

Query Match 90.8%; Score 2436; DB 1; Length 703;  
Best Local Similarity 89.7%; Pred. No. 1.6e-195;  
Matches 444; Conservative 23; Mismatches 26; Indels 2; Gaps 1;  
Qy 1 MKGMKEVLNSDAVLPSRVYVIAVOTAGYRVLRLYEGFENDASHDFWCLNLTVDVHPI 60  
Db 211 MKGMKEVLNSDAVLPSRVYVIAVOTAGYRVLRLYEGFENDASHDFWCLNLTVDVHPI 270  
Qy 61 GWCAINSKILVPPRTTHAKFTDWKYLKRLVSGRTLPVDFHIKXVSMKIPFQGMRL 120  
Db 271 GWCAINSKILVPPRTTHAKFTDWKYLKRLVSGRTLPVDFHIKXVSMKIPFQGMRL 330  
Qy 121 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGK 180  
Db 331 VVDKSOVSRTRMAVVDVTVIGRLRLLYEDGSDDDFWCHMWSPLIHPVGSRRVGHGK 390  
Qy 181 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGGMFECKMKEADPLNLGNICVATV 240  
Db 391 SERSDMAHHTPRKIYCDVPIYFKKRVAVYTGGMFECKMKEADPLNLGNICVATV 450  
Qy 241 CKVLLDGYLMICVDGGPSTGLDFWCHASSHAIFPATFCQKNDIELTPPKGYEATFNW 300  
Db 451 CKVLLDGYLMICVDGGPSTGLDFWCHASSHAIFPATFCQKNDIELTPPKGYEATFNW 510  
Qy 301 ENYLEKTSKAAPSRLEFNMDCPNHGFVKGMKLEAVDLMEPRLICVATVKKVHRLLSIHF 360  
Db 511 ENYLEKTSKAAPSRLEFNMDCPNHGFVKGMKLEAVDLMEPRLICVATVKKVHRLLSIHF 570  
Qy 361 DGWSEYDQWVDCESPDIPYVGCWELTGYQLQPPVAAPATPLKAKEATKKKKQFGKKR 420  
Db 571 DGWSEYDQWVDCESPDIPYVGCWELTGYQLQPPVAAPATPLKAKEATKKKKQFGKKR 630  
Qy 421 KRIPPTKTRPLRQSGKXPLEDDPQGARKISSEVPVGEIIAVRVKESHLDVASPDKASSP 480  
Db 631 KRIPPTKTRPLRQSGKXPLEDDPQGARKISSEVPVGEIIAVRVKESHLDVASPDKASSP 690  
Qy 481 ELPVSVENIKQETDD 495  
Db 691 ELPVSVENIKQETDD 705



RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
RA Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,  
RA Gilbert J.G.R., Goward M.B., Grafham D.V., Griffiths M.N.D., Hall C.,  
RA Hall R.E., Hall-Tanlyn G., Heathcote R.W., Ho S., Holmes S.,  
RA Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,  
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
RA Martin I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,  
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spraggon L., Steward C.A., Sulston J.E., Swann R.M.,  
RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Wiley D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.B., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
RA Minoshima S., Shibuya K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chasseo S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,  
RA Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
RA Emanuel B.S., Shaith T., Kurahashi H., Saitta S., Budarf M.L.,  
RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow S.E., Edelmann L.,  
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M.K., Kedra D.,  
RA Serousi E., Franssen I., Tapia I., Bruder C.E., O'Brien K.P.,  
RA Wilkinson P., Bodetnich A., Hartman K., Hu X., Khan A.S., Lane L.,  
RA Tilahun Y., Wright H.;  
RL "The DNA sequence of human chromosome 22.";  
RL Nature 402:489-495 (1999).  
RN (5);  
RN SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Placenta;  
RX MEDLINE=23889257; PubMed=12477932;  
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RX Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RX Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RX Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,  
RX Diatchenko L., Narusina K., Farmer A.A., Rubin G.M., Hong L.,  
RX Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
RX Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RX Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RX Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RX Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RX Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RX Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
RX Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RX Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RX Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RX Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
RX Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN (6);  
RN IDENTIFICATION OF COMPLEX WITH E2F6; TFDPI; MAX; MGA; E2F1; RNF2;  
RP RING1; RNF2; MBLR; BAT8 AND YAF2.  
RX MEDLINE=21999559; PubMed=12004135;  
RA Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.;  
RT "A complex with chromatin modifiers that occupies E2F- and  
RT Myc-responsive genes in G0 cells.";  
RL Science 296:1132-1136 (2002).  
CC -1- FUNCTION: Putative Polycomb group (PcG) protein. PcG proteins  
CC maintain the transcriptionally repressive state of genes, probably  
CC via a modification of chromatin, rendering it heritably changed in  
CC its expressibility. Its association with a chromatin remodeling

complex suggests that it may contribute to prevent expression of  
genes that trigger the cell into mitosis.  
-1- SUBUNIT: Part of the E2F6-complex in G0 phase composed of  
E2F6, MGA, MAX, TFDPI, CBX3, BAT8, E2F1, RING1, RNF2, MBLR,  
BAT8 and YAF2.  
-1- SUBCELLULAR LOCATION: Nuclear (Probable).  
-1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=3;  
Name=1; Synonyms=A;  
IsoId=Q969R5-1; Sequence=Displayed;  
Name=2; Synonyms=B;  
IsoId=Q969R5-2; Sequence=VSP\_003904; VSP\_003905;  
Name=3;  
IsoId=Q969R5-3; Sequence=VSP\_003906; VSP\_003907;  
-1- SIMILARITY: Contains 1 C4-type zinc finger.  
-1- SIMILARITY: Contains 4 MBT domains.  
-1- CAUTION: Ref.3 sequences differ from that shown in that they seem  
to include intronic sequence.  
-1- CAUTION: Ref.4 sequence differs from that shown due to erroneous  
gene model prediction.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
CC EMBL; AJ305226; CAC37794.1; -;  
CC EMBL; AJ305227; CAC37795.1; -;  
CC EMBL; AK136564; CAB66499.1; -;  
CC EMBL; AK074091; BAB84917.1; ALT SEQ.  
CC EMBL; AK097052; BAC04936.1; ALT SEQ.  
CC EMBL; AL035658; -; NOT ANNOTATED CDS.  
CC EMBL; AL035681; CAB63071.1; ALT\_SEQ.  
CC EMBL; BC017191; AAH17191.1; -;  
CC Genew; HGNC:18594; L3MBTL2.  
CC GK; Q969R5; -;  
CC InterPro; IPR004092; MBT.  
CC Pfam; PF02820; mbt; 4.  
CC SMART; SM00561; MBT; 4.  
CC Transcription regulation; Chromatin regulator; Nuclear protein;  
CC Zinc-finger; Metal-binding; Repeat; Alternative splicing;  
CC Polymorphism. 90 114 C4-TYPE.  
CC ZN\_FING 214 291 MBT 1.  
CC FT DOMAIN 327 399 MBT 2.  
CC FT DOMAIN 432 509 MBT 3.  
CC FT DOMAIN 540 611 MBT 4.  
CC FT DOMAIN 17 20 POLY-GLU.  
CC FT DOMAIN 620 624 POLY-LYS.  
CC FT VARSPPLIC 608 614 E2F6PLK -> GKLPRL (in isoform 2).  
CC FT VARSPPLIC 615 705 Missing (in isoform 2).  
CC FT VARSPPLIC 608 617 Missing (in isoform 2).  
CC FT VARSPPLIC 618 705 E2F6PLKAKE -> GVSGRGPRL (in isoform 3).  
CC FT VARSPPLIC 618 705 Missing (in isoform 3).  
CC FT VARIANT 300 300 R -> W (in dSNP:2277846).  
CC FT SEQUENCE 705 AA; 79110 MW; 8FC86A440982FFA7 CRC64;  
SQ  
Query Match 100.0%; Score 2683; DB 1; Length 705;  
Best Local Similarity 100.0%; Pred. No. 3.8e-216;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKGMKVEVLNDAVLPSRVYVIAVIQTAGVRLVYEGFENDASHDFWNCILCTVDVHPI 60  
DB 211 MKGMKVEVLNDAVLPSRVYVIAVIQTAGVRLVYEGFENDASHDFWNCILCTVDVHPI 270  
QY 61 GWCAINSKILVPPRTTHAKFTDWKGYLMKRLVSRILPVDFTHKMYESMKYPPFQGNRL 120

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 18 Seconds  
(without alignments)  
1293.234 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683  
Sequence: 1 MKGMKEVLNSDAVLPSRVY.....KASSPELPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2683	100.0	705	1 LML2_HUMAN	Q969r5 homo sapien
2	2436	90.8	703	1 LML2_MOUSE	P59178 mus musculus
3	510	19.0	772	1 LMBT_HUMAN	Q9Y468 homo sapien
4	121	4.5	867	1 SSPO_BOVIN	P98167 bos taurus
5	116.5	4.3	200	1 RL29_MYCGE	P47405 mycoplasma
6	109.5	4.1	1026	1 TAC2_HUMAN	Q95359 homo sapien
7	104	3.9	6669	1 NEBU_HUMAN	P20329 homo sapien
8	98.5	3.7	738	1 PAP_BOVIN	P25500 bos taurus
9	98.5	3.7	738	1 PAP_MOUSE	Q61183 mus musculus
10	96.5	3.6	1377	1 Z198_HUMAN	Q9ubw7 homo sapien
11	95.5	3.6	721	1 ENP1_TORCA	P14400 torpedo cal
12	92.5	3.4	479	1 CBP5_CANAL	O43101 candida alb
13	92.5	3.4	1593	1 AT12_HUMAN	P58397 homo sapien
14	91.5	3.4	245	1 H1_WAIZE	P23444 zea mays (m
15	91	3.4	401	1 EAF3_YEAST	Q12432 saccharomyc
16	91	3.4	486	1 HSI_HUMAN	P14317 homo sapien
17	90.5	3.4	2150	1 SDC3_CABEL	P34706 caenorhabdi
18	90	3.4	520	1 GAG_SIVAI	P27972 simian immu
19	90	3.4	1020	1 NFH_HUMAN	P12036 homo sapien
20	89.5	3.3	510	1 CP50_CANMA	Q12587 candida mal
21	89	3.3	3924	1 ANK2_MOUSE	Q01484 homo sapien
22	88.5	3.3	931	1 NRP2_MOUSE	O35375 mus musculus
23	88	3.3	202	1 H1_LYCPN	P40267 lycopersico
24	88	3.3	2459	1 MAPB_RAT	P15205 rattus norv
25	87.5	3.3	425	1 SMA3_HUMAN	Q92940 homo sapien
26	87.5	3.3	925	1 NRP2_RAT	Q35276 rattus norv
27	87	3.2	519	1 GAG_SIVAT	P05892 simian immu
28	86.5	3.2	307	1 VP53_BPAPS	Q9t1p5 bacterioph
29	86.5	3.2	560	1 PTK1_YEAST	P36002 saccharomyc
30	86.5	3.2	651	1 CORO_YEAST	Q06440 saccharomyc
31	86.5	3.2	1714	1 SYEP_DROME	P28668 drosophila
32	86.5	3.2	2434	1 ABC2_MOUSE	P41234 mus musculus
33	86	3.2	400	1 FDL12_MORAP	Q9Y8h5 mortierella

34	86	3.2	1156	1 JAK1_CYPCA	Q09178 cyprinus ca
35	86	3.2	1406	1 TOPI_CANGA	Q93794 candida gla
36	86	3.2	1440	1 SYEP_HUMAN	P07814 homo sapien
37	85.5	3.2	414	1 YAF4_ECOLI	P04335 escherichia
38	85.5	3.2	876	1 SYA_THEVO	Q97ab7 thermoplas
39	85	3.2	400	1 FDL3_MORIS	P59668 mortierella
40	85	3.2	484	1 PAP3_XENLA	P51005 xenopus lae
41	85	3.2	600	1 NUCD_BUCAL	P57254 buchnera ap
42	85	3.2	601	1 GLMS_PYRAE	Q8ztz0 p glucosami
43	85	3.2	619	1 PFCK_HAFEO	P29130 haemochus
44	85	3.2	1075	1 NFK3_HUMAN	Q12968 homo sapien
45	85	3.2	2468	1 MAPB_HUMAN	P46821 homo sapien

## ALIGNMENTS

RESULT 1  
LML2\_HUMAN STANDARD; PRT: 705 AA.  
AC Q969r5; QSTEN1; Q96SC4; Q9BQI2; Q9UGS4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Lethal(3)malignant brain tumor-like 2 protein (L(3)mbt-like 2 protein)  
DE (H-1(3)mbt-like protein).  
GN L3MBTL2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=21538645; PubMed=11682070;  
RA Wisnar J.;  
RT "Molecular characterization of h-l(3)mbt-like: a new member of the  
RT human mbt family.";  
RL FEBS Lett. 507:119-121(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Amnygdala;  
RC MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellerreuther R., Gassenhuber J., Glassl S.,  
RA Ansgore W., Boecker M., Bloecker H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,  
RA Meves H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
RA Wambutt R., Korn B., Klein M., Poustka A.;  
RA "Towards a catalog of human genes and proteins: sequencing and  
RA analysis of 500 novel complete protein coding human cDNAs.";  
RL Genome Res. 11:422-435(2001).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=Small intestine, and Spleen;  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,  
RA Kawakami B., Kanehori K., Sugano S., Nagahari K., Masuho Y., Nagai K.,  
RA Isogai T., Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T.,  
RA Ohara O.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,  
RA Clamp M., Smak I.J., Ainscough R., Almeida J.P., Babbage A.K.,  
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
RA Bird C.P., Blake S.E., Bridgman A.M., Buck D., Burgess J.,  
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,





QY 661 GAGCAGTCTACACAGAAAGCGGTTGGTTTGGAGAGGGATGAAGCTGGAGGCCATTGACC 720  
 Db 1295 GAGCAGTCTACACAGAAAGCGGTTGGTTTGGAGAGGGATGAAGCTGGAGGCCATTGACC 1354  
 QY 721 CCCTGAATCTGGGCAACATCTGCGTGGCAACTGCTCTGAAGGTTCTCCTGGATGGATACC 780  
 Db 1355 CCCTGAATCTGGGCAACATCTGCGTGGCAACTGCTCTGAAGGTTCTCCTGGATGGATACC 1414  
 QY 781 TGATGATCTGTGTGGACGGGGGCGCTCCACAGATGGCTTGGATGGTCTGTACATG 840  
 Db 1415 TGATGATCTGTGTGGACGGGGGCGCTCCACAGATGGCTTGGATGGTCTGTACATG 1474  
 QY 841 CCTCTTCCAGCCCATCTTCCCGGCGCCCTTCTGTGAGAGNATGACATTGAGCTCACAC 900  
 Db 1475 CCTCTTCCAGCCCATCTTCCCGGCGCCCTTCTGTGAGAGNATGACATTGAGCTCACAC 1534  
 QY 901 CGCCAAAAGGTTATGAGGACAGACTTCAACTGGGAGAACTACTTGGAGAGAACCAAGT 960  
 Db 1535 CGCCAAAAGGTTATGAGGACAGACTTCAACTGGGAGAACTACTTGGAGAGAACCAAGT 1594  
 QY 961 CGAAAGCGCTCCATCGAGACTCTTTTAACTGAGATGCGCAAACTGAGGCTTCAAGTGG 1020  
 Db 1595 CGAAAGCGCTCCATCGAGACTCTTTTAACTGAGATGCGCAAACTGAGGCTTCAAGTGG 1080  
 QY 1021 GCATGAAGCTGGAGCGGTGAGCTGATGAGAGCGCGCTCATCTGTGTGCGCAAGGTGA 1080  
 Db 1655 GCATGAAGCTGGAGCGGTGAGCTGATGAGAGCGCGCTCATCTGTGTGCGCAAGGTGA 1140  
 QY 1081 AACGAGTGTGATCGGCTCTCGAGCTTCACTTTGAGCGGTGGGACAGGAGTACGACC 1140  
 Db 1715 AACGAGTGTGATCGGCTCTCGAGCTTCACTTTGAGCGGTGGGACAGGAGTACGACC 1174  
 QY 1141 AGTGGTGGACTGGAGTCCGAGACATCTACCGGCTCGGCTGTGTGAGCTCACCGGCT 1200  
 Db 1775 AGTGGTGGACTGGAGTCCGAGACATCTACCGGCTCGGCTGTGTGAGCTCACCGGCT 1260  
 QY 1201 ACCAGCTCCAGCTCTCTGTGCGCGAGAAACCGGCGCACACCGCTCAAGCGCCAAAGAGGCCA 1260  
 Db 1835 ACCAGCTCCAGCTCTCTGTGCGCGAGAAACCGGCGCACACCGCTCAAGCGCCAAAGAGGCCA 1320  
 QY 1261 CAAGAAGAAAGAAACAGTTTGGAGAGAAAGAAAGAAATCCCGGCCACTAAGACGC 1320  
 Db 1895 CAAGAAGAAAGAAACAGTTTGGAGAGAAAGAAAGAAATCCCGGCCACTAAGACGC 1380  
 QY 1321 GACCCCTCAGACAGGGGTCCAGAGAGCCCTGTGAGAGAGACCGCTCAGGGTCCAGGA 1380  
 Db 1955 GACCCCTCAGACAGGGGTCCAGAGAGCCCTGTGAGAGAGACCGCTCAGGGTCCAGGA 2014  
 QY 1381 AGATCTCGTGGAGCGCTGTCTCTGCGAGATCATGTGTGCGTGTGAAGAGAGCATC 1440  
 Db 2015 AGATCTCGTGGAGCGCTGTCTCTGCGAGATCATGTGTGCGTGTGAAGAGAGCATC 2074  
 QY 1441 TAGAGTGGCTGCGCGACAGAGGTTCAAGTCCAGAGCTGCTGCTCCGTCAGAGACA 1500  
 Db 2075 TAGAGTGGCTGCGCGACAGAGGTTCAAGTCCAGAGCTGCTGCTCCGTCAGAGACA 2134  
 QY 1501 TCAAGCAGGAAACAGACGACTGAGCGCTTCTGCGCTCCAGCGCTGCTTCTAGCTGGAAGCC 1560  
 Db 2135 TCAAGCAGGAAACAGACGACTGAGCGCTTCTGCGCTCCAGCGCTGCTTCTAGCTGGAAGCC 2194  
 QY 1561 AGCCAGCGGTTTCTTACACCAACCAAGCTTCCAGCTGAGCTTGGCTGGAGACTGA 1620  
 Db 2195 AGCCAGCGGTTTCTTACACCAACCAAGCTTCCAGCTGAGCTTGGCTGGAGACTGA 2254  
 QY 1621 TCCTCTGTGTAAATCTTCTGCGCGGTGTGTGAAGGCTGGAGCGTGGAGCGCTGCGG 1680  
 Db 2255 TCCTCTGTGTAAATCTTCTGCGCGGTGTGTGAAGGCTGGAGCGTGGAGCGCTGCGG 2314  
 QY 1681 GTCTCTGTGGAACCGGCTGTGCTTCTGCGCTTCCCTTGGAAAGGTTATATGACGGGC 1740  
 Db 2315 GTCTCTGTGGAACCGGCTGTGCTTCTGCGCTTCCCTTGGAAAGGTTATATGACGGGC 2374  
 QY 1741 CGCCTGAGGCGCCAGAACTGCTGCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGCTG 1800

Db 2375 CGCCTGAGGCCCCAGAACTGCTCTGTGAACCACTTTTCAGCCAGAGTTCCCAAGCTG 2434  
 QY 1801 GAAAGCTAGCTGCTGCTTCTTCTTAAAGATGGCTCCCGCCGACCCCGCCAGCGGCTCAG 1860  
 Db 2435 GAAAGCTAGCTGCTGCTTCTTCTTAAAGATGGCTCCCGCCGACCCCGCCAGCGGCTCAG 2494  
 QY 1861 TTGCCAGGGATGGGGCCACACTGTCACACTGTGGAAATACAGACAGTGAATCTGTCTG 1920  
 Db 2495 TTGCCAGGGATGGGGCCACACTGTCACACTGTGGAAATACAGACAGTGAATCTGTCTG 2554  
 QY 1921 CTTGAACAGAGTCATGTAAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATGAAGTCCCT 1980  
 Db 2555 CTTGAACAGAGTCATGTAAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATGAAGTCCCT 2614  
 QY 1981 GACAGTGAATGTGTGGGGGCGAGCTCTGCTTCAAAAATTCACAGAGAGATGCT 2040  
 Db 2615 GACAGTGAATGTGTGGGGGCGAGCTCTGCTTCAAAAATTCACAGAGAGATGCT 2674  
 QY 2041 CTTGAGCTCATGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2100  
 Db 2675 CTTGAGCTCATGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2734  
 QY 2101 CTTGCTCGGAGCTTAAGAACAGTGAACAGATGTGATTTTGGCGACCTGTGTGTGG 2160  
 Db 2735 CTTGCTCGGAGCTTAAGAACAGTGAACAGATGTGATTTTGGCGACCTGTGTGTGG 2794  
 QY 2161 CTTGAGCTCATGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2220  
 Db 2795 CTTGAGCTCATGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2854  
 QY 2221 GGGGAGGACATTTGGGAGGAAAGTGGCTGAGTGTGCTCTGCTCTGCTCTGCTCTGCTCT 2280  
 Db 2855 GGGGAGGACATTTGGGAGGAAAGTGGCTGAGTGTGCTCTGCTCTGCTCTGCTCTGCTCT 2914  
 QY 2281 GAAGCCCGCTTAAATAATTAATCAATCAAGATTCCTTTGTAGTAAAGGTCCAGTTCTGA 2340  
 Db 2915 GAAGCCCGCTTAAATAATTAATCAATCAAGATTCCTTTGTAGTAAAGGTCCAGTTCTGA 2974  
 QY 2341 CTTGAGCTCTAGAGCTGGGCTGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 2400  
 Db 2975 CTTGAGCTCTAGAGCTGGGCTGTGTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 3034  
 QY 2401 AAGAAACCAATGCTGGAGGGCGGTGAACACAGAACCTCAAGACAGATGACAGAGCT 2460  
 Db 3035 AAGAAACCAATGCTGGAGGGCGGTGAACACAGAACCTCAAGACAGATGACAGAGCT 3094  
 QY 2461 GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCGAGCTCTGTGTGTGAGA 2520  
 Db 3095 GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCGAGCTCTGTGTGTGAGA 3154  
 QY 2521 AATTAAACCCCTGCTTGTGCTTGAAGAAAAA 2555  
 Db 3155 AATTAAACCCCTGCTTGTGCTTGAAGAAAAA 3189

RESULT 3  
 BC017191  
 LOCUS  
 DEFINITION  
 Homo sapiens, hypothetical protein DKFZp761i1141, clone MGC:2476  
 IMAGE:3138444, mRNA, complete cds.  
 ACCESSION  
 VERSION  
 BC017191  
 KEYWORDS  
 MGC.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE  
 1 (bases 1 to 3222)  
 AUTHORS  
 Strausberg, R.  
 TITLE  
 Direct Submission  
 JOURNAL  
 Submitted (05-NOV-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 18 Seconds  
(without alignments)  
1293.234 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKVELNSDAVLPSRVY.....KASPELPSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2683	100.0	705	1 LML2_HUMAN	Q969r5 homo sapien
2	2436	90.8	703	1 LML2_MOUSE	P59178 mus musculus
3	510	19.0	772	1 LMBT_HUMAN	Q9Y468 homo sapien
4	121	4.5	867	1 SSP0_BOVIN	P98167 bos taurus
5	116.5	4.3	200	1 RL29_MYCCE	P47405 mycoplasma
6	109.5	4.1	1026	1 TAC2_HUMAN	O95359 homo sapien
7	104	3.9	6869	1 NEBU_HUMAN	P20529 homo sapien
8	98.5	3.7	738	1 PAP_BOVIN	P25500 bos taurus
9	98.5	3.7	738	1 PAP_MOUSE	Q61183 mus musculus
10	96.5	3.6	1377	1 Z198_HUMAN	Q9ubw7 homo sapien
11	95.5	3.6	721	1 ENP1_TORCA	P14400 torpedo cal
12	92.5	3.4	479	1 CBP5_CANAL	O43101 candida alb
13	92.5	3.4	1593	1 ATL2_HUMAN	P58397 homo sapien
14	91.5	3.4	245	1 H1_MAIZE	P23444 zea mays (m
15	91	3.4	401	1 EAF3_YEAST	Q12432 saccharomyc
16	91	3.4	486	1 HSI_HUMAN	P14317 homo sapien
17	90.5	3.4	2150	1 SD03_CAEEL	P34706 caenorhabdi
18	90	3.4	520	1 GAG_SIVAI	P27972 simian immu
19	90	3.4	1020	1 NEH_HUMAN	P12036 homo sapien
20	89.5	3.3	510	1 CP53_CANMA	Q2587 candida mal
21	89	3.3	3924	1 ANK2_HUMAN	Q01484 homo sapien
22	88.5	3.3	931	1 NRP2_MOUSE	O35375 mus musculus
23	88	3.3	202	1 H1_LYCPN	P40267 lycopersico
24	88	3.3	2459	1 MAPB_RAT	P15205 rattus norv
25	87.5	3.3	425	1 SMA3_HUMAN	Q92940 homo sapien
26	87.5	3.3	925	1 NRP2_RAT	O35276 rattus norv
27	87	3.2	519	1 GAG_SIVAT	P05892 simian immu
28	86.5	3.2	307	1 VP53_EPAFS	Q9t1p5 bacterioph
29	86.5	3.2	560	1 PTK1_YEAST	P36002 saccharomyc
30	86.5	3.2	651	1 CORO_YEAST	Q06440 saccharomyc
31	86.5	3.2	1714	1 SYEP_DROME	P28668 drosophila
32	86.5	3.2	2434	1 ABC2_MOUSE	P41234 mus musculus
33	86	3.2	400	1 FD12_MORAP	Q9y8h5 mortierella

RESULT 1  
LML2\_HUMAN  
ID LML2\_HUMAN STANDARD; PRT; 705 AA.  
AC Q969r5; Q8TEN1; Q96SC4; Q9BQI2; Q9UGS4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Lethal(3)malignant brain tumor-like 2 protein (L(3)mbt-like 2 protein)  
DE (H-1(3)mbt-like protein).  
GN L3MBTL2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE=21538645; PubMed=11682070;  
RA Wismar J.;  
RT "Molecular characterization of h-1(3)mbt-like: a new member of the  
human mbt family.";  
RL FEBS Lett. 507:119-121(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=amygdala;  
RX MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glaesli S.,  
RA Ansoerge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
RA Lauber J., Dueterhoeft A., Beyer A., Koshner K., Strack N.,  
RA Mewes H.-W., Oosterwaelder B., Obermaier B., Tampe J., Heubner D.,  
RA Wambutt K., Korn B., Klein M., Poustka A.;  
RT "Towards a catalog of human genes and proteins: sequencing and  
analysis of 500 novel complete protein coding human cDNAs.";  
Genome Res. 11:422-435(2001).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 3).  
RC TISSUE=small intestine, and spleen;  
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,  
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,  
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,  
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,  
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,  
RA Murakawa K., Nakahori K., Takahashi-Fujii A., Oshino A., Sugiyama A.,  
RA Kawakami K., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,  
RA Isogai T., Jikuya H., Takano J., Nomura N., Kikuno R., Nagase T.,  
RA Ohara O.;  
RT "NEDO human cDNA sequencing project.";  
RP Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20057165; PubMed=10591208;  
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,  
RA Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,  
RA Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,  
RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,  
RA Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,  
RA Clegg S.M., Cooley V.E., Cole C.G., Collier R.E., Connor R.,

#### ALIGNMENTS

RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,  
RA Dhani P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,  
RA Evans K.L., Fey J.C., Fleming K., French L., Garner A.A.,  
RA Gilbert J.G.R., Goward M.E., Graffham D.V., Griffiths M.N.D., Hall C.,  
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,  
RA Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,  
RA Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,  
RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,  
RA Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,  
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,  
RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,  
RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,  
RA Soderlund C., Spragon L., Steward C.A., Suleston J.E., Swann R.N.,  
RA Vaubin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,  
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,  
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,  
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,  
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,  
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,  
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Leo H.I.,  
RA Lewis J., Lewis S., Lin S.-P., Loh P., Mala J.E., Nguyen T., Pan H.,  
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,  
RA Wang Q., Wang Y., Chang S., White J., Willingham D., Wu H., Yao Z.,  
RA Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,  
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,  
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,  
RA Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Kohlfing T.,  
RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,  
RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,  
RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,  
RA McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,  
RA Kim U.J., Shizuya H., Simon M.I., Dumanaki J.P., Peyrard M., Kedra D.,  
RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P., Lane L.,  
RA Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,  
RA Tilahun Y., Wright H.;  
RT "The DNA sequence of human chromosome 22.";  
RL Nature 402:489-495(1999).  
[5]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE2Placenta;  
RX MEDLINE=2389257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).  
[6]  
RN IDENTIFICATION OF COMPLEX WITH E2F6; TFDPI; MAX; MGA; EHMUTASE1; CBX3;  
RP RING1; RNF2; MBLR; BAT8 AND YAP2.  
RX MEDLINE=2199559; PubMed=12004135;  
RA Ogawa H., Ishiguro K.-I., Gaubatz S., Livingston D.M., Nakatani Y.;  
RT "A complex with chromatin modifiers that occupies E2F- and  
Myc-responsive genes in G0 cells.";  
RL Science 296:1132-1136(2002).  
CC -!- FUNCTION: Putative Polycomb group (PcG) protein. PcG proteins  
maintain the transcriptionally repressive state of genes, probably  
via a modification of chromatin, rendering it heritably changed in  
its expressibility. Its association with a chromatin remodeling

complex suggests that it may contribute to prevent expression of  
genes that trigger the cell into mitosis.  
-!- SUBUNIT: Part of the E2F6.com-1 complex composed of  
E2F6, MGA, MAX, TFDPI, CBX3, BAT8, EHMUTASE1, RING1, RNF2, MBLR,  
BAT8 and YAP2.  
-!- SUBCELLULAR LOCATION: Nuclear (Probable).  
-!- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=3;  
Name=1; Synonyms=A;  
IsoId=Q969R5-1; Sequences=Displayed;  
Name=2; Synonyms=B;  
IsoId=Q969R5-2; Sequences=VSP\_003904, VSP\_003905;  
Name=3;  
IsoId=Q969R5-3; Sequences=VSP\_003906, VSP\_003907;  
-!- SIMILARITY: Contains 1 C4-type zinc finger.  
-!- SIMILARITY: Contains 4 MBL domains.  
-!- CAUTION: Ref.3 sequences differ from that shown in that they seem  
to include intronic sequence.  
-!- CAUTION: Ref.4 sequence differs from that shown due to erroneous  
gene model prediction.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).  
-----  
EMBL; AJ305226; CAC37794.1; -;  
EMBL; AJ305227; CAC37795.1; -;  
EMBL; AL136564; CAB66499.1; -;  
EMBL; AK074091; BAB84917.1; ALT\_SEQ.  
EMBL; AK097052; BAC04936.1; ALT\_SEQ.  
EMBL; AL035658; -; NOT\_ANNOTATED\_CDS.  
EMBL; AL035681; CAB63071.1; ALT\_SEQ.  
EMBL; BC017191; AAH17191.1; -;  
Genew; HGNC:18594; L3MBTL2.  
GK; Q969R5; -;  
InterPro; IPR004092; Mbt.  
Pfam; PF02820; mbt; 4.  
SMART; SM00561; MBL; 4.  
Transcription regulation; Chromatin regulator; Nuclear protein;  
Zinc-finger; Metal-binding; Repeat; Alternative splicing;  
Polymorphism.  
Zn\_Fing 90 114 C4-TYPE.  
DOMAIN 214 291 MET 1.  
DOMAIN 327 399 MET 2.  
DOMAIN 432 509 MET 3.  
DOMAIN 540 611 MET 4.  
DOMAIN 17 20 POLY-LYS.  
DOMAIN 620 624 POLY-LYS.  
VARSPIC 608 614 EPATPK -> GKLPRESL (in isoform 2).  
VARSPIC 615 705 Missing (in isoform 2).  
VARSPIC 608 617 EPATPKAKE -> GVGSRGPKRL (in isoform 3).  
VARSPIC 618 705 Missing (in isoform 3).  
VARIANT 300 300 R -> W (IN dbSNP:2277846).  
SEQ SEQUENCE 705 AA; 79110 MW; 8FC86A440982FFA7 CRC64;  
Query Match 100.0%; Score 2683; DB 1; Length 705;  
Best Local Similarity 100.0%; Pred. No. 3.8e-216;  
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 MKGKVEVLNSDAVLPSPRYTWIASVQTAGYRVLLRYEGFENDASHDFWCLGTVDVHP1 60  
Db 211 MKGKVEVLNSDAVLPSPRYTWIASVQTAGYRVLLRYEGFENDASHDFWCLGTVDVHP1 270  
Qy 61 GWCAINSKILVPPRTTHAKFTDVKYLMKRLVGSRTLPVDFIKMVESMKYPPFQGMRL2 120



```

RESULT 3
ID LMST HUMAN STANDARD; PRT; 772 AA.
AC Q9Y468; Q9H1E5; Q9UG05; Q9UB99; Q9Y4C9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Lethal(3)malignant brain tumor-like protein (L(3)mbl-like) (L(3)mbl
GN L3MBTL OR L3MBT OR KIA0681.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=93373015; PubMed=10445843;
RA Koga H., Matsui S.-I., Hirota T., Takebayashi S.-I., Okumura K.,
RA Sava H.;
RT "A human homolog of Drosophila lethal(3)malignant brain tumor
RT (L(3)mbl) protein associates with condensed mitotic chromosomes.";
RL Oncogene 18:3799-3809(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Uterus;
RX Kohrer K., Beyer A., Mewes H.-W., Cassenhuber J., Wiemann S.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burdill W.D., Butler A.B., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaeslatho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [4]
RP SEQUENCE OF 215-772 FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:159-176(1998).
CC -1- FUNCTION: Polycomb group (PcG) protein. PcG proteins maintain the
CC transcriptionally repressive state of genes, probably via a
CC modification of chromatin, rendering it heritably changed in its
CC expressibility. Probably plays a role in cell proliferation.
CC Overexpression induces multinucleated cells, suggesting that it is

```

```

required to accomplish normal mitosis.
-1- SUBCELLULAR LOCATION: Nuclear; excluded from the nucleolus. Does
not colocalizes with the PcG protein BMI1, suggesting that these
two proteins do not belong to the same complex.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=4;
Name=1; Synonyms=mbt-1;
IsoId=Q9Y468-1; Sequence=Displayed;
Name=2; Synonyms=mbt-11;
IsoId=Q9Y468-2; Sequence=VSP_003902;
Name=3;
IsoId=Q9Y468-3; Sequence=VSP_003901; VSP_003902;
Name=4;
IsoId=Q9Y468-4; Sequence=VSP_003903;
-1- TISSUE SPECIFICITY: Widely expressed. Expression is reduced in
colorectal cancer cell line SW480 and promyelocytic leukemia cell
line HL-60.
-1- DEVELOPMENTAL STAGE: In interphase cells, it is scattered
throughout the nucleoplasm. In mitotic cells, it strongly
associates with condensed chromosomes from the prophase to
telophase.
-1- SIMILARITY: Contains 3 mbt domains.
-1- CAUTION: Ref.3 (CAC18508) sequences differ from that shown due to
erroneous gene model prediction.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; U89358; AAC69438.1; --
EMBL; AL110279; CAB3714.1; --
EMBL; Z98752; CAC16799.1; --
EMBL; Z98752; CAC16800.1; --
EMBL; Z98752; CAC18508.1; --
EMBL; AL031681; CAC17518.1; ALT_SEQ.
EMBL; AL031681; CAB43959.1; --
EMBL; AB014581; BAA31656.1; --
FIR; T14794; T14794.
Gene; HGNC:15905; L3MBTL.
GK; Q9Y468; --
InterPro; IPR004092; Mbt.
InterPro; IPR002515; ZnF_C2HC.
Pfam; PF02820; mbt; 3.
Pfam; PF01530; ZF-C2HC; 1.
SMART; SM00561; MBT; 3.
Transcription regulation; Chromatin regulator; Zinc-finger;
DNA-binding; Nuclear protein; Repeat; Alternative splicing.
DOMAIN 242 315 MBT 1.
DOMAIN 349 422 MBT 2.
DOMAIN 453 526 MBT 3.
ZNFING 552 578 C2HC-TYPE.
VARSPLIC 1 348 Missing (in isoform 3).
VARSPLIC 709 772 /FTid=VSP_003901.
ARIVRVTHVSGKTLVTVTAQGLDVCSDHLOEGKILETV
HSLICSLPHTLLAKLSPASDSQ -> VRCKRVGDRAGVT
VLKTAGRCPCQRHFC (in isoform 2 and
isoform 3).
/FTid=VSP_003902.
ARIVRVTHVSGKTLVTVTAQGLDVCSDHLOEGKILETV
HSLICSLPHTLLAKLSPASDSQ -> MIDGEAPLLLTQAD
IVKIMSVKLGPKLYNAILMFKVADDTLK (in
isoform 4).
/FTid=VSP_003903.
P -> L (IN REF. 1).
LR -> MC (IN REF. 1).
L -> M (IN REF. 1).
S -> P (IN REF. 1).
SEQUENCE 772 AA; 85916 MW; 117B03A628826B29 CRC64;

```

Query Match		19.0%; Score 510; DB 1; Length 772;
Best Local Similarity		33.7%; Pred. No. 1.2e-34;
Matches 136; Conservative		53; Mismatches 168; Indels 47; Gaps 14;
QY	83	WGYLMKRLVGRSLTPVDF--HIKAVESMKYFFRQGMRLVVDKQSVSRTRMAVVDTVIG 140
Db	208	WESYLEEQ--KAITAPVSLFQDSQAVTHNKGKFLGMKLEGIDPQHPMSMYILITVAEYCG 265
QY	141	GLRLLYEDGSDDDDFCHWMSPLTHPVGWSRVRVGHGIMSE--RRSDMAHHPTFRKIYC 198
Db	266	YLRHLHFDGYSCHDFWVANSFDPHPAGWFKTKLQPPKGYKEEFWSQVLRSTRA 325
QY	199	DAVP--YLFKKVRAVTEGGWFEKMLRAIDPLNIGNICVATCKVLLDGLVLMICVDGCP 257
Db	326	QAAPKHLFVSQSHSPPLG-FQVGMKLEAVDEMNSLVCVASVTDV-VDSFLVHFDNWD 383
QY	258	STDGLDWFYHASSHAIPATPCQKNDIELTPPKGY-EAOTFNMYNLEYLTKSKAAPSRL 316
Db	384	DT--YDYWC-DFSSPIIHPVGMQCGKPLTPPDYDPDNFCWEKYLEETGASAVPTWA 440
QY	317	FNNDCPNHGFVGMKLEAVDLMEPLICVATKRVVHLLSIHFDGWDSEYDQWDCESP 376
Db	441	FKVR--PPHSFLVMKLEAVDRNPALIRVASVEDVHRKIHFHDSHGSHYDFWIDADHP 499
QY	377	DIYPVGCWELTGY-QPPVA-AEPAT-----PLKAEATKXKKKQFGKKRKP----- 424
Db	500	DIHPAGWCSKTGHPLOPLGPREPSSASPGGCPPLSYRSLPHTRTSKYSEHHRKCPGPG 559
QY	425	-----PKTRPLRQSKKPLEDDPPQARK 449
Db	560	DOSGHVTKFTAHNCLSGCPAER--NQSRLKAELSSEASARK 601
RESULT 4		
ID	SSPO BOVIN	STANDARD; PRT; 867 AA.
AC	P98167;	
DI	01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1996 (Rel. 34, Last sequence update)	
DT	28-FEB-2003 (Rel. 41, Last annotation update)	
DE	SCO-spondin (Fragment)	
OS	Bos taurus (Bovine)	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;	
OC	Bovidae; Bovinae; Bos.	
OX	NCBI_TaxID=9913;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE-Endothocyte;	
RX	MEDLINE=98338614; PubMed=8743952;	
RA	Gobron S., Monnerie H., Meinzel R., Creveaux I., Lehmann W.,	
RA	Lamalle D., Pastugue B., Meinzel A.;	
RT	"SCO-spondin: a new member of the thrombospondin family secreted by	
RT	the subcommissural organ is a candidate in the modulation of neuronal	
RT	aggregation.";	
RL	J Cell Sci. 109:1053-1061(1996).	
CC	- FUNCTION: INVOLVED IN THE MODULATION OF NEURONAL AGGREGATION.	
CC	- SUBCELLULAR LOCATION: Extracellular.	
CC	- TISSUE SPECIFICITY: SUBCOMMISSURAL ORGAN.	
CC	- DEVELOPMENTAL STAGE: EMBRYO.	
CC	- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.	
CC	- SIMILARITY: Contains at least 4 TSP type-1 domains.	
CC	- SIMILARITY: Contains at least 2 EGF-like domains.	
CC	- SIMILARITY: Contains at least 1 FS/8 type C domain.	
CC	- SIMILARITY: Contains at least 3 LDL-receptor class A domains.	
CC	- SIMILARITY: Contains 1 VWF domain.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )	
		-----
or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> .		
CC	EMBL; X93922; CAA63815.1; -	
DR	HSP; P01130; IAUJ.	
DR	InterPro; IPR000421; FA58 C.	
DR	InterPro; IPR002172; LDL_Receptor_A.	
DR	InterPro; IPR002913; TIL_Cybrich.	
DR	InterPro; IPR000884; TSP1.	
DR	InterPro; IPR001007; VWF C.	
DR	Pfam; PF00754; F5_F8_type_C; 1.	
DR	Pfam; PF00057; ldl_recept_a; 3.	
DR	Pfam; PF01826; TSP_1; 4.	
DR	Pfam; PF00030; TSP_1; 4.	
DR	Pfam; PF00033; VWF; 1.	
DR	SMART; SM00331; FA58C; 1.	
DR	SMART; SM00192; LDLA; 3.	
DR	SMART; SM00209; TSP1; 4.	
DR	SMART; SM00214; VWC; 1.	
DR	PROSITE; PS01285; FA58C_1; 1.	
DR	PROSITE; PS01286; FA58C_2; 1.	
DR	PROSITE; PS00022; FA58C_3; 1.	
DR	PROSITE; PS01209; LDLRA_1; 3.	
DR	PROSITE; PS00068; LDLRA_2; 3.	
DR	PROSITE; PS00092; TSP1; 4.	
DR	PROSITE; PS0184; VWF; 2; 1.	
KW	Glycoprotein; Cell adhesion; Calcium-binding; Repeat; EGF-like domain.	
FT	NON_TER	1
FT	DOMAIN	29 87
FT	DOMAIN	103 142
FT	DOMAIN	143 180
FT	DOMAIN	188 244
FT	DOMAIN	245 304
FT	DOMAIN	344 502
FT	DOMAIN	506 544
FT	DOMAIN	563 701
FT	DOMAIN	723 761
FT	DOMAIN	816 866
FT	DISULFID	107 122
FT	DISULFID	116 127
FT	DISULFID	129 141
FT	DISULFID	147 166
FT	DISULFID	149 169
FT	DISULFID	171 179
FT	DISULFID	344 502
FT	DISULFID	508 520
FT	DISULFID	535 533
FT	DISULFID	527 542
FT	DISULFID	665 677
FT	DISULFID	672 690
FT	DISULFID	694 699
FT	DISULFID	725 737
FT	DISULFID	732 750
FT	DISULFID	744 759
FT	CARBOHYD	88 88
FT	CARBOHYD	309 309
FT	CARBOHYD	409 409
FT	NON_TER	867 867
SQ	SEQUENCE	867 AA; 91817 MW; 9538F2108E787B49 CRC64;
Query Match		4.5%; Score 121; DB 1; Length 867;
Best Local Similarity		20.8%; Pred. No. 0.041;
Matches 75; Conservative		38; Mismatches 130; Indels 118; Gaps 17;
QY	161	WSPLIHPV-----GWSRRVGHGKMSERRSDMAHPTFRKIYCDVAPVYLFKKVAVYTEGG 216
Db	370	WAAILRPAPGAGWS-----PVEHADTQGHPP-----PVL----- 400
QY	217	WFEENKLEADPLNIGNICVATCKVLLDGLVLMICVDGSPSTGLDWFCHASSHA--- 273
Db	401	-----QLDLQPRNLGTIIYQAGS-----SDWLQV-----SSDGLHWSYRDIGHGTQP 445
QY	274	---IFF-----ATFCQKNDIBLTTPPKGYEAQTFNWNYLENTEKSKAAPSRLFN 318



```
Db 446 APQLFPKNGKSTVWMPFARMVQARHVRWPSDGH-----QAAPSSDAN 490
QY 319 MDCENHGFVGKGLKLEAVDLMEPRILCVATKRVVHLLSIHF-----DGWD 364
Db 491 LDGP-----LRVELLG-CEPAPICLG-----VGRVCVSGCAPGAPCDGVECKDQSD 538
QY 365 SEYDOWDCEPDYIFVGCETLQYLOPPVAAEPATPLKAKEATKKKKQFGK--KKR 422
Db 539 EE-----GCVTPPA-GAGRIESTAWSAPSSAQPCQLPPQPSSEGLAEADHMHGPGSP 592
QY 423 LPPTKTRPLRGSKFLEDDPQARKISSEPVGEIIVAVKHEHLDVADSPDKASSPEL 482
Db 593 VPPTKKGASLGSEB---HPSPGSGVQTVTPTSQPEAQALRPEVAATVLPHPHMTVEV 649
QY 483 P 483
Db 650 P 650

RESULT 5
RL29_MYCGB
ID RL29_MYCGB STANDARD; PRT; 200 AA.
AC P47405;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L29.
GN RPMC OR RPL29 OR MG159.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium."
RL Science 270:397-403(1995).
CC -!- SIMILARITY: BELONGS TO THE L29P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; U39696; AAC71377.1; -
CC FIR; F64217; F64217.
CC TIGR; MG159; -.
CC HAMAP; MF_00374; fused; 1.
CC InterPro; IPR001854; Ribosomal L29.
CC Pfam; PF00831; Ribosomal L29; 1.
CC TIGRFAM; TIGR00012; L29; 1.
CC PROSITE; PS00579; RIBOSOMAL_L29; 1.
KW Ribosomal protein. Complete proteome.
FT DOMAIN 1 77 50S RIBOSOMAL PROTEIN L29.
FT DOMAIN 78 200 UNKNOWN.
SQ SEQUENCE 200 AA; 23258 MW; 5BC115E52AA5EA92 CRC64;

Query Match 4.3%; Score 116.5; DB 1; Length 200;
Best Local Similarity 25.0%; Pred. No. 0.014; Indels 29; Gaps 6;
Matches 44; Conservative 38; Mismatches 65;

QY 331 KLEAVDLMEPRILCVATKRVVHLLSIHFD---QWDSEYDQW--VDCESPDYFVGW-- 383
```

```
Db 32 KLAHGELDKPHLI--AKVRKLLAVLVTLITERKLANVQVEKQYKLLSRKTNELVNSMKQ 89
QY 384 -----CELTGYLOPPVAAEPATPLKAKEATKKKKQFGKRRKIPPTKTRPLRQG 434
Db 90 KUSTPESKQETTKAEVFKVESKESKQETTKAEVKKLQETKKVEVKVPKPEPLKQE 149
QY 435 SKKPLEDDPQARKISSEPVGEIIVAVKHEHLDVADSPDKASSPELPVSVENIK 490
Db 150 TKK--VEARIETKTKVESKPLKQEVKVEAKK-----SVSKPQKPVKARMIK 194

RESULT 6
TACC2_HUMAN
ID TACC2_HUMAN STANDARD; PRT; 1026 AA.
AC O95359; Q9NZ41; Q9NZRS;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transforming acidic coiled-coil-containing protein 2 (Anti Zuai-1)
DE (AZU-1).
GN TACC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain, Fetal brain, and skeletal muscle;
RX MEDLINE=20570483; PubMed=11121038;
RA Gergely F., Karlsson C., Still I.H., Cowell J.K., Kilmartin J.,
RA Raff J.W.;
RT "The TACC domain identifies a family of centrosomal proteins that can
RT interact with microtubules."
RL Proc. Natl. Acad. Sci. U.S.A. 97:14352-14357(2000).
RN [2]
RP SEQUENCE OF 375-1026 FROM N.A. (ISOFORM 2).
RX MEDLINE=20214826; PubMed=10749935;
RA Chen H.-M., Schmeichel K.L., Mian I.S., Lelievre S., Petersen O.W.,
RA Bissell M.J.;
RT "AZU-1: a candidate breast tumor suppressor and biomarker for tumor
RT progression."
RL Mol. Biol. Cell 11:1357-1367(2000).
RN [3]
RP SEQUENCE OF 44-1026 FROM N.A. (ISOFORM 3).
RA Pu J., Li C., Rodriguez M., Banerjee D.;
RT "Expression of TACC2 protein mRNA in human microvascular endothelial
RT cells."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY PLAY A ROLE IN ORGANIZING CENTROSOMAL MICROTUBULES.
CC MAY ACT AS A TUMOR SUPPRESSOR PROTEIN. MAY REPRESENT A TUMOR
CC PROGRESSION MARKER.
CC -!- SUBUNIT: INTERACTS WITH MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR; CONCENTRATED AT CENTROSOMES.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Experimental confirmation may be lacking for some
CC isoforms;
CC Name=1;
CC IsoId=O95359-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O95359-2; Sequence=VSP_006368; VSP_006369;
CC Name=3;
CC IsoId=O95359-3; Sequence=VSP_006369;
CC -!- SIMILARITY: BELONGS TO THE TACC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
```

```
DR EMBL; AF095791; AAC64968.2; --
DR EMBL; AF176646; AAF63433.1; ALT_INIT.
DR EMBL; AF220152; AAF29537.2; --
DR Genew; HGNC:11523; TACC2.
DR MIM; 605302; --
DR GO; GO:0007048; P:oncogenesis; TAS.
DR Pfam; PF05010; TACC; 1.
KW Coiled coil; Nuclear protein; Alternative splicing.
FT DOMAIN 43 51 POLY-PRO.
FT DOMAIN 498 501 POLY-LYS.
FT DOMAIN 753 781 COILED COIL (POTENTIAL).
FT DOMAIN 824 1025 COILED COIL (POTENTIAL).
FT VARSPPLIC 507 510 Missing (in isoform 2).
FT VARSPPLIC 711 787 Missing (in isoform 2 and isoform 3).
FT VARSPPLIC 711 787 Missing (in isoform 2 and isoform 3).
FT CONFLICT 44 67 PAPPPPPPEVPEVSTQPPPEE -> MFWYKRRGAPMRP
FT VSVIDGVVCS (IN REF. 3).
SQ SEQUENCE 1026 AA; 112110 MW; E2575FCB446E9CF8 CRC64;

Query Match 4.1%; Score 109.5; DB 1; Length 1026;
Best Local Similarity 23.6%; Pred. NO. 0.47;
Matches 57; Conservative 33; Mismatches 95; Indels 57; Gaps 11;

QY 253 VDGPGSTGLDFWFCVHASSHALEPATFCQKNDIELTPKGYEAOTFNWENYLEKTKSKAA 312
DQ 77 VPDGPRSDVSGSPFRPPSHS-FAVFDKPIASSGTNYNLDPNIELVDTFTQLEPRAS 135
QY 313 PSRLFNMDCPNHFQKVGKMLKLEAVLMEPRLLICVATKRVVHRLLSIH---FDGWDSDYDQ 369
DQ 136 -----DAKQEQKVNTRKSTD-----SVISKSTLSRLSLQASDFDQASSSGNP 181
QY 370 WVDCESPDIYPVGCETGYQLPPVAAEPATPLKAKEATKK-----KKQFGKKKKRI 423
DQ 182 EAVLAPAPAYSTG-----SSASSTLK---RTKKPRPPSLKKQITTK--- 221
QY 424 PPTKTRPLRQSKKPLLEDPPQGARKISSEVPV-GEIIAVRVKKEHLDVASPDKASSPEL 482
DQ 222 -PTETPPVKYEQEP-----DRESL-----VPSGENLASSTKTESAKTEGSPALLEET 269
QY 483 PV 484
DQ 270 PL 271

RESULT 7
NEBU_HUMAN STANDARD; PRT; 6669 AA.
AC P20929; Q15346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nebulin.
GN NEB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. PubMed=7739042;
RX MEDLINE=95257391; PubMed=7739042;
RA Labeit S., Kolmerer B.;
RT "The complete primary structure of human nebulin and its correlation
to muscle structure.";
RL J. Mol. Biol. 248:308-315 (1995).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE.
RX MEDLINE=88284704; PubMed=3397062;
RA Zeviani M., Darras B.T., Rizzuto R., Salviati G., Betto R.,
RA Bonilla E., Miranda A.F., Du J., Samitt C., Dickson G., Walsh P.S.,
RA Dimauro S., Francke U., Schon E.A.;
RT "Cloning and expression of human nebulin cDNAs and assignment of the
gene to chromosome 2q31-q32";
```

```
Genomics 2:249-256 (1988).
[3]
STRUCTURE BY NMR OF 6610-6669.
MEDLINE=98179559; PubMed=9514727;
RA Politou A.S., Millevoi S., Gautel M., Kolmerer B., Pastore A.;
RT "SH3 in muscles: solution structure of the SH3 domain from nebulin.";
J. Mol. Biol. 276:189-202 (1998).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
OF STRIATED MUSCLE.
CC -1- DISEASE: Defects in NEB are a cause of the autosomal recessive
form of nemaline myopathy (NEM2).
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -1- SIMILARITY: Contains 178 nebulin repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X83957; CAAS8788.1; --
EMBL; M19668; AAAS9916.1; ALT_SEQ.
EMBL; M19669; AAAS9917.1; ALT_SEQ.
DR PIR; S55024; S55024.
DR PDB; 1ARK; 28-JAN-98.
DR PDB; 1NEB; 24-DEC-97.
DR Genew; HGNC:7720; NEB.
DR MIM; 161650; --
DR GO; GO:0015629; C:actin cytoskeleton; TAS.
DR GO; GO:0030017; C:sarcomere; NAS.
DR GO; GO:0003792; F:regulation of actin thin filament length ac. .; NAS.
DR GO; GO:0008307; F:structural constituent of muscle; TAS.
DR GO; GO:0007525; P:somatic muscle development; NAS.
DR InterPro; IPR000900; Nebulin.
DR Pfam; PF00880; Nebulin; 146.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PRO0510; NEBULIN.
DR PRINTS; PRO0452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 76 107 NEBULIN 1.
FT REPEAT 113 143 NEBULIN 2.
FT REPEAT 148 178 NEBULIN 3.
FT REPEAT 183 213 NEBULIN 4.
FT REPEAT 218 248 NEBULIN 5.
FT REPEAT 253 283 NEBULIN 6.
FT REPEAT 289 318 NEBULIN 7.
FT REPEAT 324 354 NEBULIN 8.
FT REPEAT 363 393 NEBULIN 9.
FT REPEAT 398 428 NEBULIN 10.
FT REPEAT 434 454 NEBULIN 11.
FT REPEAT 502 532 NEBULIN 12.
FT REPEAT 537 567 NEBULIN 13.
FT REPEAT 573 603 NEBULIN 14.
FT REPEAT 611 641 NEBULIN 15.
FT REPEAT 681 711 NEBULIN 16.
FT REPEAT 749 779 NEBULIN 17.
FT REPEAT 784 814 NEBULIN 18.
FT REPEAT 820 850 NEBULIN 19.
FT REPEAT 858 888 NEBULIN 20.
FT REPEAT 893 923 NEBULIN 21.
FT REPEAT 924 954 NEBULIN 22.
FT REPEAT 959 990 NEBULIN 23.
FT REPEAT 993 1023 NEBULIN 24.
FT REPEAT 1028 1058 NEBULIN 25.
```

FT	REPEAT	3603	3633	NEBULIN 99.
FT	REPEAT	3638	3668	NEBULIN 100.
FT	REPEAT	3671	3701	NEBULIN 101.
FT	REPEAT	3706	3736	NEBULIN 102.
FT	REPEAT	3742	3772	NEBULIN 103.
FT	REPEAT	3780	3810	NEBULIN 104.
FT	REPEAT	3815	3845	NEBULIN 105.
FT	REPEAT	3846	3876	NEBULIN 106.
FT	REPEAT	3914	3944	NEBULIN 107.
FT	REPEAT	3949	3979	NEBULIN 108.
FT	REPEAT	3984	4014	NEBULIN 109.
FT	REPEAT	4021	4052	NEBULIN 110.
FT	REPEAT	4057	4087	NEBULIN 111.
FT	REPEAT	4088	4118	NEBULIN 112.
FT	REPEAT	4123	4153	NEBULIN 113.
FT	REPEAT	4156	4186	NEBULIN 114.
FT	REPEAT	4191	4220	NEBULIN 115.
FT	REPEAT	4226	4256	NEBULIN 116.
FT	REPEAT	4264	4294	NEBULIN 117.
FT	REPEAT	4299	4329	NEBULIN 118.
FT	REPEAT	4330	4360	NEBULIN 119.
FT	REPEAT	4365	4395	NEBULIN 120.
FT	REPEAT	4400	4430	NEBULIN 121.
FT	REPEAT	4435	4465	NEBULIN 122.
FT	REPEAT	4471	4501	NEBULIN 123.
FT	REPEAT	4544	4574	NEBULIN 124.
FT	REPEAT	4575	4605	NEBULIN 125.
FT	REPEAT	4610	4640	NEBULIN 126.

Query Match

Best Local Similarity

Mismatches

Conservative

Score 104; DB 1; Length 6669;

19.6%; Pred. No. 17;

54; Mismatches 100; Indels 142; Gaps 19;

QY	160	MWSPLIHPVGSRRVGHG-----IKMSRRSDMA-----HHPTFRKI	196	
Db	2266	LYSQKLYKLGWEALKKGYDLPVDAISVQLAKASRDIAADYKQGYKQKLGHHVGFSL	2325	
QY	197	YCDAPVYLFKVRVYVTEGGWFE--EGMKLEAIDPLNLGNCVATVCKVLLD-----GYL	249	
Db	2326	ODDPKLVLSMNVAKMQSREYKKDFEKKTKPSSPVDMLGVVLAKKCCQELSDVDYKNYL	2385	
QY	250	MICVDGPGSTDGLDWFYCHASHAIFPATFCQKNDIELTPPKGYEAQTFN-----	299	
Db	2386	H-----QWTC-----LPQNDV-VQAKKVYELQSENLYKSDLEWLR	2420	
QY	300	--WE--NYLEKTKSKAA-----PSRLEFNMDCPNHGFVKVGMKLEAVDLMEPRL	342	
Db	2421	GIGWSPLGSLAEKKNKGRASEIISEKKYRQPPDRNKFTSIP-----DAMD	2465	
QY	343	ICVATVVRVHRLSIHPDGDWSEYDQ-WVDCESPD	-----YFVWCBL-	386
Db	2466	--VLAKTNAKRSRLYREAWDKTKQHIMEDTDPDIVLAKANLINTSDKLYRMGYSELK	2523	
QY	387	-TGYYLQPPVAAEPATPLKAKEATKKKKQKQKGRKRIPTKTRPLRQSGKKPL-----	439	
Db	2524	RKGYDL--PV---DAIPKAKASREIASEYKK-----EGFRQLGHGICA	2565	
QY	440	--LEDDPQ	445	
Db	2566	RNIEDDPK	2573	

RESULT 8

PAP\_BOVIN

ID

PAP\_BOVIN

STANDARD;

PRT;

738 AA.

AC

P25500;

DT

01-MAY-1992 (Rel. 22, Created)

DT

01-FEB-1994 (Rel. 28, Last sequence update)

DT

28-FEB-2003 (Rel. 41, Last annotation update)

DE

Poly(A) polymerase alpha (EC 2.7.7.19) (PAP) (Polynucleotide

DE

adenylityltransferase alpha).

GN

PAP01A OR PAP.

OS

Bos taurus (Bovine).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]\_SEQUENCE FROM N.A., AND SEQUENCE OF 1-20; 206-254 AND 385-396.  
RP TISSUE=thymus;  
RC MEDLINE=92097545; PubMed=1756732;  
RA Wahle E., Martin G., Schiltz E., Keller W.;  
RT "Isolation and expression of cDNA clones encoding mammalian poly(A)  
RT polymerase.";  
RL EMBO J. 10:4251-4257(1991).  
RN [2]  
RP REVISIONS.  
RA Wahle E.;  
RL Submitted (AUG-1992) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 96-102 AND 444-467.  
RC TISSUE=Heart muscle;  
EX MEDLINE=91375535; PubMed=1896071;  
RA Raabe T., Bollum F.J., Manley J.L.;  
RT "Primary structure and expression of bovine poly(A) polymerase.";  
RL Nature 353:229-234(1991).  
RN [4]  
RP DOMAINS, ACTIVE SITES, AND MUTAGENESIS.  
EX MEDLINE=96221320; PubMed=8665867;  
RA Martin G., Keller W.;  
RT "Mutational analysis of mammalian poly(A) polymerase identifies a  
RT region for primer binding and catalytic domain, homologous to the  
RT family X polymerases, and to other nucleotidyltransferases.";  
RL EMBO J. 15:2593-2603(1996).  
RN [5]  
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).  
RX MEDLINE=20403225; PubMed=1094102;  
RA Martin G., Keller W., Double S.;  
RT "Crystal structure of mammalian poly(A) polymerase in complex with an  
RT analog of ATP.";  
RL EMBO J. 19:4193-4203(2000).  
CC -!- FUNCTION: Polymerase that creates the 3' poly(A) tail of mRNA's.  
CC Also required for the endoribonucleolytic cleavage reaction at  
CC some polyadenylation sites. May acquire specificity through  
CC interaction with a cleavage and polyadenylation specificity factor  
CC (CPSF) at its C-terminus.  
CC -!- CATALYTIC ACTIVITY: N ATP + [nucleotide] (M) = N diphosphate +  
CC [nucleotide] (M+N).  
CC -!- SUBUNIT: Monomer.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=Long;  
CC IsoId=P25500-1; Sequence=Displayed;  
CC Name=Short;  
CC IsoId=P25500-2; Sequence=VSP\_004524, VSP\_004525, VSP\_004526;  
CC Note=No experimental confirmation available;  
CC -!- PTM: Phosphorylated. Phosphorylation/dephosphorylation may  
CC regulate the interaction between PAP and CPSF.  
CC -!- SIMILARITY: BELONGS TO THE POLY(A) POLYMERASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X61585; CAA43782.1; -;  
CC EMBL; X63436; CAA45031.1; -;  
CC PDB; 1F5A; 13-SEP-00.  
DR InterPro; IPR002934; NTP transf.  
DR InterPro; IPR001201; PAP\_25A core.  
DR InterPro; IPR007012; PAP\_cent.  
  
DR InterPro; IPR007010; PAP\_RNA\_bind.  
DR Pfam; PF01909; NTP transf\_2; 1.  
DR Pfam; PF04928; PAP\_central; 1.  
DR Pfam; PF04926; PAP\_RNA\_bind; 1.  
KW mRNA processing; Transferase; Transcription; RNA-binding;  
KW Nuclear protein; Phosphorylation; Alternative splicing; 3D-structure.  
FT INIT\_MET 0 0  
FT ACT\_SITE 112 112  
FT ACT\_SITE 114 114  
FT ACT\_SITE 166 166  
FT DOMAIN 489 506  
FT DOMAIN 643 658  
FT VARSPIC 662 682  
FT VARSPIC 708 709  
FT VARSPIC 710 738  
FT CONFLICT 79 79  
FT HELIX 32 45  
FT HELIX 46 48  
FT TURN 49 49  
FT HELIX 54 81  
FT TURN 82 83  
FT HELIX 86 89  
FT TURN 90 91  
FT STRAND 95 98  
FT HELIX 100 104  
FT TURN 105 105  
FT TURN 109 110  
FT STRAND 113 119  
FT TURN 121 122  
FT HELIX 125 128  
FT TURN 129 130  
FT HELIX 131 137  
FT TURN 138 138  
FT TURN 140 141  
FT STRAND 142 148  
FT TURN 149 150  
FT STRAND 155 160  
FT TURN 161 162  
FT STRAND 163 171  
FT TURN 179 180  
FT TURN 183 184  
FT HELIX 186 188  
FT TURN 189 191  
FT HELIX 194 211  
FT TURN 212 212  
FT HELIX 216 232  
FT TURN 233 234  
FT STRAND 236 236  
FT TURN 238 241  
FT STRAND 243 243  
FT HELIX 245 258  
FT TURN 260 261  
FT HELIX 264 276  
FT TURN 277 277  
FT TURN 280 281  
FT STRAND 284 284  
FT TURN 301 303  
FT HELIX 305 309  
FT STRAND 314 315  
FT STRAND 323 323  
FT TURN 324 327  
FT HELIX 330 351  
FT TURN 352 353  
FT HELIX 357 359  
FT TURN 360 361  
FT HELIX 366 369  
FT STRAND 372 379  
FT HELIX 383 394  
FT TURN 395 396  
FT HELIX 397 405  
  
InterPro; IPR007010; PAP\_RNA\_bind.  
Pfam; PF01909; NTP transf\_2; 1.  
Pfam; PF04928; PAP\_central; 1.  
Pfam; PF04926; PAP\_RNA\_bind; 1.  
mRNA processing; Transferase; Transcription; RNA-binding;  
Nuclear protein; Phosphorylation; Alternative splicing; 3D-structure.  
INIT\_MET 0 0  
ACT\_SITE 112 112  
ACT\_SITE 114 114  
ACT\_SITE 166 166  
DOMAIN 489 506  
DOMAIN 643 658  
VARSPIC 662 682  
VARSPIC 708 709  
VARSPIC 710 738  
CONFLICT 79 79  
HELIX 32 45  
HELIX 46 48  
TURN 49 49  
HELIX 54 81  
TURN 82 83  
HELIX 86 89  
TURN 90 91  
STRAND 95 98  
HELIX 100 104  
TURN 105 105  
TURN 109 110  
STRAND 113 119  
TURN 121 122  
HELIX 125 128  
TURN 129 130  
HELIX 131 137  
TURN 138 138  
TURN 140 141  
STRAND 142 148  
TURN 149 150  
STRAND 155 160  
TURN 161 162  
STRAND 163 171  
TURN 179 180  
TURN 183 184  
HELIX 186 188  
TURN 189 191  
HELIX 194 211  
TURN 212 212  
HELIX 216 232  
TURN 233 234  
STRAND 236 236  
TURN 238 241  
STRAND 243 243  
HELIX 245 258  
TURN 260 261  
HELIX 264 276  
TURN 277 277  
TURN 280 281  
STRAND 284 284  
TURN 301 303  
HELIX 305 309  
STRAND 314 315  
STRAND 323 323  
TURN 324 327  
HELIX 330 351  
TURN 352 353  
HELIX 357 359  
TURN 360 361  
HELIX 366 369  
STRAND 372 379  
HELIX 383 394  
TURN 395 396  
HELIX 397 405  
  
NUCLEAR LOCALIZATION SIGNAL 1.  
NUCLEAR LOCALIZATION SIGNAL 2.  
Missing (in isoform Short).  
/FTId=VSP\_004524.  
KT -> II (in isoform Short).  
/FTId=VSP\_004525.  
Missing (in isoform Short).  
/FTId=VSP\_004526.  
S -> R (IN REF. 3).

```

FT TURN 406 406
FT TURN 408 409
FT STRAND 410 415
FT STRAND 420 421
FT STRAND 433 442
FT HELIX 456 472
FT TURN 473 474
FT TURN 478 479
FT STRAND 481 488
FT HELIX 489 491
FT HELIX 493 495
SQ SEQUENCE 738 AA; 82310 MW; DOB90662F89363E0 CRC64;

Query Match 3.7%; Score 98.5; DB 1; Length 738;
Best Local Similarity 20.9%; Fred. No. 2.5;
Matches 53; Conservative 33; Mismatches 105; Indels 63; Gaps 9;

Qy 274 IFPATFCQKN-DIELTPPKGYEAQTFNWNYLEKTKSKAAPSRLFNMDCPNHGFKVGMKL 332
Db 441 VFKEKTENSENLVDLT---YDIQSFTDYYRQAINSKM-----FEYDMKI 482

Qy 333 EAVDLMEPRILCVATKRVVHRLLSIH-----DGEYDQWDCESP 377
Db 483 AAMHVKRKQL-----HQLPSHLVQKKKHSTGVKLTPLNDSSLDLSMDSNDNM 532

Qy 378 IYPVGWCELTCYOLQ-----PPVAABPATPLKAKEAT---KKKKQFGKKK 421
Db 533 SVSPSTAMKTSPLNSGSGQGNPAPAVTAASVTNIQTEVSLPQINSSESGGTSS 592

Qy 422 RIPPTKTRPQSGKFLLEDDPQGARKISSEVPGEIIAVRVKEHLVAVSPDKASSPE 481
Db 593 SIPTATQPAISPPKPTSVRSVSTLVNPPRPSCNAAKIPN---PIVGKRTSSPH 649

Qy 482 LPVSVENIKQETDD 495
Db 650 KESPPKTKTEDE 663

RESULT 9
PAP_MOUSE
ID_ PAP_MOUSE STANDARD; PRT; 738 AA.
AC Q61183; Q61208; Q61209; Q8K4X2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Poly(A) polymerase alpha (EC 2.7.7.19) (PAP) (Polynucleotide
DE adenylyltransferase)
GN PAPOLA OR PAP OR PLAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS 2; 3 AND 4).
RC TISSUE=Brain;
RX MEDLINE=96189120; PubMed=8628305;
RA Zhao W., Manley J.L.;
RT "Complex alternative RNA processing generates an unexpected diversity
RL Mol. Cell. Biol. 16:2378-2386(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=129/SvJ;
RX MEDLINE=22359775; PubMed=12471261;
RA Kashiwabara S., Noguchi J., Zhuang T., Ohmura K., Honda A., Baba T.;
RA Sugitara S., Miyamoto K., Takahashi S., Inoue K., Ogura A.;
RT "Regulation of spermatogenesis by testis-specific, cytoplasmic
RL Poly(A) polymerase TPAP";
RL Science 298:1999-2002(2002).
CC -/- FUNCTION: Polymerase that creates the 3' poly(A) tail of mRNA's.
CC Also required for the endoribonucleolytic cleavage reaction at
CC some polyadenylation sites. May acquire specificity through
CC interaction with a cleavage and polyadenylation specificity factor

```

```

CC (CPSF) at its C-terminus (By similarity).
CC -/- CATALYTIC ACTIVITY: N ATP + {nucleotide}(M) = N diphosphate +
CC {nucleotide}(M+N).
CC -/- SUBUNIT: Monomer (By similarity).
CC -/- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -/- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1;
CC IsoId=Q61183-1; Sequence=Displayed;
CC Name=2; Synonyms=III;
CC IsoId=Q61183-2; Sequence=VSP_004531, VSP_004532;
CC Name=3; Synonyms=V;
CC IsoId=Q61183-3; Sequence=VSP_004529, VSP_004530;
CC Name=4; Synonyms=VI;
CC IsoId=Q61183-4; Sequence=VSP_004527, VSP_004528;
CC -/- PTM: Phosphorylated. Phosphorylation/dephosphorylation may
CC regulate the interaction between PAP and CPSF (By similarity).
CC -/- SIMILARITY: BELONGS TO THE POLY(A) POLYMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U52197; AAC52586.1; -.
CC EMBL; U58134; AAC52608.1; -.
CC EMBL; U58135; AAC52609.1; -.
CC EMBL; AB086650; BAC00996.1; -.
CC EMBL; AB086633; BAC00996.1; JOINED.
CC EMBL; AB086634; BAC00996.1; JOINED.
CC EMBL; AB086635; BAC00996.1; JOINED.
CC EMBL; AB086636; BAC00996.1; JOINED.
CC EMBL; AB086637; BAC00996.1; JOINED.
CC EMBL; AB086638; BAC00996.1; JOINED.
CC EMBL; AB086639; BAC00996.1; JOINED.
CC EMBL; AB086640; BAC00996.1; JOINED.
CC EMBL; AB086641; BAC00996.1; JOINED.
CC EMBL; AB086642; BAC00996.1; JOINED.
CC EMBL; AB086643; BAC00996.1; JOINED.
CC EMBL; AB086644; BAC00996.1; JOINED.
CC EMBL; AB086645; BAC00996.1; JOINED.
CC EMBL; AB086646; BAC00996.1; JOINED.
CC EMBL; AB086647; BAC00996.1; JOINED.
CC EMBL; AB086648; BAC00996.1; JOINED.
CC EMBL; AB086649; BAC00996.1; JOINED.
CC HSP; P25500; 1F5A.
CC MGD; MGI:109301; Papola.
CC InterPro; IPR002934; NTP transf.
CC InterPro; IPR001201; PAP_25A core.
CC InterPro; IPR007012; PAP-centr.
CC InterPro; IPR007010; PAP-RNA-bind.
CC Pfam; PF04926; PAP-RNA-bind; 1.
CC mRNA processing; Transferase; Transcription; RNA-binding;
CC Phosphorylation; Nuclear protein; Alternative splicing.
CC INIT MET 0 0 BY SIMILARITY.
CC ACT_SITE 112 112 BY SIMILARITY.
CC ACT_SITE 114 114 BY SIMILARITY.
CC ACT_SITE 166 166 BY SIMILARITY.
CC VARSPPLIC 279 292 EWPNFVLKQFEC -> YVFRLYNKKIDCRH (in
CC isoform 4).
CC /FtId=VSP_004527.
CC /FtId=VSP_004528.
CC /FtId=VSP_004529.
CC /FtId=VSP_004530.
CC /FtId=VSP_004531.
CC /FtId=VSP_004532.
CC /FtId=VSP_004533.
CC /FtId=VSP_004534.
CC /FtId=VSP_004535.
CC /FtId=VSP_004536.
CC /FtId=VSP_004537.
CC /FtId=VSP_004538.
CC /FtId=VSP_004539.
CC /FtId=VSP_004540.
CC /FtId=VSP_004541.
CC /FtId=VSP_004542.
CC /FtId=VSP_004543.
CC /FtId=VSP_004544.
CC /FtId=VSP_004545.
CC /FtId=VSP_004546.
CC /FtId=VSP_004547.
CC /FtId=VSP_004548.
CC /FtId=VSP_004549.
CC /FtId=VSP_004550.
CC /FtId=VSP_004551.
CC /FtId=VSP_004552.
CC /FtId=VSP_004553.
CC /FtId=VSP_004554.
CC /FtId=VSP_004555.
CC /FtId=VSP_004556.
CC /FtId=VSP_004557.
CC /FtId=VSP_004558.
CC /FtId=VSP_004559.
CC /FtId=VSP_004560.
CC /FtId=VSP_004561.
CC /FtId=VSP_004562.
CC /FtId=VSP_004563.
CC /FtId=VSP_004564.
CC /FtId=VSP_004565.
CC /FtId=VSP_004566.
CC /FtId=VSP_004567.
CC /FtId=VSP_004568.
CC /FtId=VSP_004569.
CC /FtId=VSP_004570.
CC /FtId=VSP_004571.
CC /FtId=VSP_004572.
CC /FtId=VSP_004573.
CC /FtId=VSP_004574.
CC /FtId=VSP_004575.
CC /FtId=VSP_004576.
CC /FtId=VSP_004577.
CC /FtId=VSP_004578.
CC /FtId=VSP_004579.
CC /FtId=VSP_004580.
CC /FtId=VSP_004581.
CC /FtId=VSP_004582.
CC /FtId=VSP_004583.
CC /FtId=VSP_004584.
CC /FtId=VSP_004585.
CC /FtId=VSP_004586.
CC /FtId=VSP_004587.
CC /FtId=VSP_004588.
CC /FtId=VSP_004589.
CC /FtId=VSP_004590.
CC /FtId=VSP_004591.
CC /FtId=VSP_004592.
CC /FtId=VSP_004593.
CC /FtId=VSP_004594.
CC /FtId=VSP_004595.
CC /FtId=VSP_004596.
CC /FtId=VSP_004597.
CC /FtId=VSP_004598.
CC /FtId=VSP_004599.
CC /FtId=VSP_004600.
CC /FtId=VSP_004601.
CC /FtId=VSP_004602.
CC /FtId=VSP_004603.
CC /FtId=VSP_004604.
CC /FtId=VSP_004605.
CC /FtId=VSP_004606.
CC /FtId=VSP_004607.
CC /FtId=VSP_004608.
CC /FtId=VSP_004609.
CC /FtId=VSP_004610.
CC /FtId=VSP_004611.
CC /FtId=VSP_004612.
CC /FtId=VSP_004613.
CC /FtId=VSP_004614.
CC /FtId=VSP_004615.
CC /FtId=VSP_004616.
CC /FtId=VSP_004617.
CC /FtId=VSP_004618.
CC /FtId=VSP_004619.
CC /FtId=VSP_004620.
CC /FtId=VSP_004621.
CC /FtId=VSP_004622.
CC /FtId=VSP_004623.
CC /FtId=VSP_004624.
CC /FtId=VSP_004625.
CC /FtId=VSP_004626.
CC /FtId=VSP_004627.
CC /FtId=VSP_004628.
CC /FtId=VSP_004629.
CC /FtId=VSP_004630.
CC /FtId=VSP_004631.
CC /FtId=VSP_004632.
CC /FtId=VSP_004633.
CC /FtId=VSP_004634.
CC /FtId=VSP_004635.
CC /FtId=VSP_004636.
CC /FtId=VSP_004637.
CC /FtId=VSP_004638.
CC /FtId=VSP_004639.
CC /FtId=VSP_004640.
CC /FtId=VSP_004641.
CC /FtId=VSP_004642.
CC /FtId=VSP_004643.
CC /FtId=VSP_004644.
CC /FtId=VSP_004645.
CC /FtId=VSP_004646.
CC /FtId=VSP_004647.
CC /FtId=VSP_004648.
CC /FtId=VSP_004649.
CC /FtId=VSP_004650.
CC /FtId=VSP_004651.
CC /FtId=VSP_004652.
CC /FtId=VSP_004653.
CC /FtId=VSP_004654.
CC /FtId=VSP_004655.
CC /FtId=VSP_004656.
CC /FtId=VSP_004657.
CC /FtId=VSP_004658.
CC /FtId=VSP_004659.
CC /FtId=VSP_004660.
CC /FtId=VSP_004661.
CC /FtId=VSP_004662.
CC /FtId=VSP_004663.
CC /FtId=VSP_004664.
CC /FtId=VSP_004665.
CC /FtId=VSP_004666.
CC /FtId=VSP_004667.
CC /FtId=VSP_004668.
CC /FtId=VSP_004669.
CC /FtId=VSP_004670.
CC /FtId=VSP_004671.
CC /FtId=VSP_004672.
CC /FtId=VSP_004673.
CC /FtId=VSP_004674.
CC /FtId=VSP_004675.
CC /FtId=VSP_004676.
CC /FtId=VSP_004677.
CC /FtId=VSP_004678.
CC /FtId=VSP_004679.
CC /FtId=VSP_004680.
CC /FtId=VSP_004681.
CC /FtId=VSP_004682.
CC /FtId=VSP_004683.
CC /FtId=VSP_004684.
CC /FtId=VSP_004685.
CC /FtId=VSP_004686.
CC /FtId=VSP_004687.
CC /FtId=VSP_004688.
CC /FtId=VSP_004689.
CC /FtId=VSP_004690.
CC /FtId=VSP_004691.
CC /FtId=VSP_004692.
CC /FtId=VSP_004693.
CC /FtId=VSP_004694.
CC /FtId=VSP_004695.
CC /FtId=VSP_004696.
CC /FtId=VSP_004697.
CC /FtId=VSP_004698.
CC /FtId=VSP_004699.
CC /FtId=VSP_004700.
CC /FtId=VSP_004701.
CC /FtId=VSP_004702.
CC /FtId=VSP_004703.
CC /FtId=VSP_004704.
CC /FtId=VSP_004705.
CC /FtId=VSP_004706.
CC /FtId=VSP_004707.
CC /FtId=VSP_004708.
CC /FtId=VSP_004709.
CC /FtId=VSP_004710.
CC /FtId=VSP_004711.
CC /FtId=VSP_004712.
CC /FtId=VSP_004713.
CC /FtId=VSP_004714.
CC /FtId=VSP_004715.
CC /FtId=VSP_004716.
CC /FtId=VSP_004717.
CC /FtId=VSP_004718.
CC /FtId=VSP_004719.
CC /FtId=VSP_004720.
CC /FtId=VSP_004721.
CC /FtId=VSP_004722.
CC /FtId=VSP_004723.
CC /FtId=VSP_004724.
CC /FtId=VSP_004725.
CC /FtId=VSP_004726.
CC /FtId=VSP_004727.
CC /FtId=VSP_004728.
CC /FtId=VSP_004729.
CC /FtId=VSP_004730.
CC /FtId=VSP_004731.
CC /FtId=VSP_004732.
CC /FtId=VSP_004733.
CC /FtId=VSP_004734.
CC /FtId=VSP_004735.
CC /FtId=VSP_004736.
CC /FtId=VSP_004737.
CC /FtId=VSP_004738.
CC /FtId=VSP_004739.
CC /FtId=VSP_004740.
CC /FtId=VSP_004741.
CC /FtId=VSP_004742.
CC /FtId=VSP_004743.
CC /FtId=VSP_004744.
CC /FtId=VSP_004745.
CC /FtId=VSP_004746.
CC /FtId=VSP_004747.
CC /FtId=VSP_004748.
CC /FtId=VSP_004749.
CC /FtId=VSP_004750.
CC /FtId=VSP_004751.
CC /FtId=VSP_004752.
CC /FtId=VSP_004753.
CC /FtId=VSP_004754.
CC /FtId=VSP_004755.
CC /FtId=VSP_004756.
CC /FtId=VSP_004757.
CC /FtId=VSP_004758.
CC /FtId=VSP_004759.
CC /FtId=VSP_004760.
CC /FtId=VSP_004761.
CC /FtId=VSP_004762.
CC /FtId=VSP_004763.
CC /FtId=VSP_004764.
CC /FtId=VSP_004765.
CC /FtId=VSP_004766.
CC /FtId=VSP_004767.
CC /FtId=VSP_004768.
CC /FtId=VSP_004769.
CC /FtId=VSP_004770.
CC /FtId=VSP_004771.
CC /FtId=VSP_004772.
CC /FtId=VSP_004773.
CC /FtId=VSP_004774.
CC /FtId=VSP_004775.
CC /FtId=VSP_004776.
CC /FtId=VSP_004777.
CC /FtId=VSP_004778.
CC /FtId=VSP_004779.
CC /FtId=VSP_004780.
CC /FtId=VSP_004781.
CC /FtId=VSP_004782.
CC /FtId=VSP_004783.
CC /FtId=VSP_004784.
CC /FtId=VSP_004785.
CC /FtId=VSP_004786.
CC /FtId=VSP_004787.
CC /FtId=VSP_004788.
CC /FtId=VSP_004789.
CC /FtId=VSP_004790.
CC /FtId=VSP_004791.
CC /FtId=VSP_004792.
CC /FtId=VSP_004793.
CC /FtId=VSP_004794.
CC /FtId=VSP_004795.
CC /FtId=VSP_004796.
CC /FtId=VSP_004797.
CC /FtId=VSP_004798.
CC /FtId=VSP_004799.
CC /FtId=VSP_004800.
CC /FtId=VSP_004801.
CC /FtId=VSP_004802.
CC /FtId=VSP_004803.
CC /FtId=VSP_004804.
CC /FtId=VSP_004805.
CC /FtId=VSP_004806.
CC /FtId=VSP_004807.
CC /FtId=VSP_004808.
CC /FtId=VSP_004809.
CC /FtId=VSP_004810.
CC /FtId=VSP_004811.
CC /FtId=VSP_004812.
CC /FtId=VSP_004813.
CC /FtId=VSP_004814.
CC /FtId=VSP_004815.
CC /FtId=VSP_004816.
CC /FtId=VSP_004817.
CC /FtId=VSP_004818.
CC /FtId=VSP_004819.
CC /FtId=VSP_004820.
CC /FtId=VSP_004821.
CC /FtId=VSP_004822.
CC /FtId=VSP_004823.
CC /FtId=VSP_004824.
CC /FtId=VSP_004825.
CC /FtId=VSP_004826.
CC /FtId=VSP_004827.
CC /FtId=VSP_004828.
CC /FtId=VSP_004829.
CC /FtId=VSP_004830.
CC /FtId=VSP_004831.
CC /FtId=VSP_004832.
CC /FtId=VSP_004833.
CC /FtId=VSP_004834.
CC /FtId=VSP_004835.
CC /FtId=VSP_004836.
CC /FtId=VSP_004837.
CC /FtId=VSP_004838.
CC /FtId=VSP_004839.
CC /FtId=VSP_004840.
CC /FtId=VSP_004841.
CC /FtId=VSP_004842.
CC /FtId=VSP_004843.
CC /FtId=VSP_004844.
CC /FtId=VSP_004845.
CC /FtId=VSP_004846.
CC /FtId=VSP_004847.
CC /FtId=VSP_004848.
CC /FtId=VSP_004849.
CC /FtId=VSP_004850.
CC /FtId=VSP_004851.
CC /FtId=VSP_004852.
CC /FtId=VSP_004853.
CC /FtId=VSP_004854.
CC /FtId=VSP_004855.
CC /FtId=VSP_004856.
CC /FtId=VSP_004857.
CC /FtId=VSP_004858.
CC /FtId=VSP_004859.
CC /FtId=VSP_004860.
CC /FtId=VSP_004861.
CC /FtId=VSP_004862.
CC /FtId=VSP_004863.
CC /FtId=VSP_004864.
CC /FtId=VSP_004865.
CC /FtId=VSP_004866.
CC /FtId=VSP_004867.
CC /FtId=VSP_004868.
CC /FtId=VSP_004869.
CC /FtId=VSP_004870.
CC /FtId=VSP_004871.
CC /FtId=VSP_004872.
CC /FtId=VSP_004873.
CC /FtId=VSP_004874.
CC /FtId=VSP_004875.
CC /FtId=VSP_004876.
CC /FtId=VSP_004877.
CC /FtId=VSP_004878.
CC /FtId=VSP_004879.
CC /FtId=VSP_004880.
CC /FtId=VSP_004881.
CC /FtId=VSP_004882.
CC /FtId=VSP_004883.
CC /FtId=VSP_004884.
CC /FtId=VSP_004885.
CC /FtId=VSP_004886.
CC /FtId=VSP_004887.
CC /FtId=VSP_004888.
CC /FtId=VSP_004889.
CC /FtId=VSP_004890.
CC /FtId=VSP_004891.
CC /FtId=VSP_004892.
CC /FtId=VSP_004893.
CC /FtId=VSP_004894.
CC /FtId=VSP_004895.
CC /FtId=VSP_004896.
CC /FtId=VSP_004897.
CC /FtId=VSP_004898.
CC /FtId=VSP_004899.
CC /FtId=VSP_004900.
CC /FtId=VSP_004901.
CC /FtId=VSP_004902.
CC /FtId=VSP_004903.
CC /FtId=VSP_004904.
CC /FtId=VSP_004905.
CC /FtId=VSP_004906.
CC /FtId=VSP_004907.
CC /FtId=VSP_004908.
CC /FtId=VSP_004909.
CC /FtId=VSP_004910.
CC /FtId=VSP_004911.
CC /FtId=VSP_004912.
CC /FtId=VSP_004913.
CC /FtId=VSP_004914.
CC /FtId=VSP_004915.
CC /FtId=VSP_004916.
CC /FtId=VSP_004917.
CC /FtId=VSP_004918.
CC /FtId=VSP_004919.
CC /FtId=VSP_004920.
CC /FtId=VSP_004921.
CC /FtId=VSP_004922.
CC /FtId=VSP_004923.
CC /FtId=VSP_004924.
CC /FtId=VSP_004925.
CC /FtId=VSP_004926.
CC /FtId=VSP_004927.
CC /FtId=VSP_004928.
CC /FtId=VSP_004929.
CC /FtId=VSP_004930.
CC /FtId=VSP_004931.
CC /FtId=VSP_004932.
CC /FtId=VSP_004933.
CC /FtId=VSP_004934.
CC /FtId=VSP_004935.
CC /FtId=VSP_004936.
CC /FtId=VSP_004937.
CC /FtId=VSP_004938.
CC /FtId=VSP_004939.
CC /FtId=VSP_004940.
CC /FtId=VSP_004941.
CC /FtId=VSP_004942.
CC /FtId=VSP_004943.
CC /FtId=VSP_004944.
CC /FtId=VSP_004945.
CC /FtId=VSP_004946.
CC /FtId=VSP_004947.
CC /FtId=VSP_004948.
CC /FtId=VSP_004949.
CC /FtId=VSP_004950.
CC /FtId=VSP_004951.
CC /FtId=VSP_004952.
CC /FtId=VSP_004953.
CC /FtId=VSP_004954.
CC /FtId=VSP_004955.
CC /FtId=VSP_004956.
CC /FtId=VSP_004957.
CC /FtId=VSP_004958.
CC /FtId=VSP_004959.
CC /FtId=VSP_004960.
CC /FtId=VSP_004961.
CC /FtId=VSP_004962.
CC /FtId=VSP_004963.
CC /FtId=VSP_004964.
CC /FtId=VSP_004965.
CC /FtId=VSP_004966.
CC /FtId=VSP_004967.
CC /FtId=VSP_004968.
CC /FtId=VSP_004969.
CC /FtId=VSP_004970.
CC /FtId=VSP_004971.
CC /FtId=VSP_004972.
CC /FtId=VSP_004973.
CC /FtId=VSP_004974.
CC /FtId=VSP_004975.
CC /FtId=VSP_004976.
CC /FtId=VSP_004977.
CC /FtId=VSP_004978.
CC /FtId=VSP_004979.
CC /FtId=VSP_004980.
CC /FtId=VSP_004981.
CC /FtId=VSP_004982.
CC /FtId=VSP_004983.
CC /FtId=VSP_004984.
CC /FtId=VSP_004985.
CC /FtId=VSP_004986.
CC /FtId=VSP_004987.
CC /FtId=VSP_004988.
CC /FtId=VSP_004989.
CC /FtId=VSP_004990.
CC /FtId=VSP_004991.
CC /FtId=VSP_004992.
CC /FtId=VSP_004993.
CC /FtId=VSP_004994.
CC /FtId=VSP_004995.
CC /FtId=VSP_004996.
CC /FtId=VSP_004997.
CC /FtId=VSP_004998.
CC /FtId=VSP_004999.
CC /FtId=VSP_005000.
CC /FtId=VSP_005001.
CC /FtId=VSP_005002.
CC /FtId=VSP_005003.
CC /FtId=VSP_005004.
CC /FtId=VSP_005005.
CC /FtId=VSP_005006.
CC /FtId=VSP_005007.
CC /FtId=VSP_005008.
CC /FtId=VSP_005009.
CC /FtId=VSP_005010.
CC /FtId=VSP_005011.
CC /FtId=VSP_005012.
CC /FtId=VSP_005013.
CC /FtId=VSP_005014.
CC /FtId=VSP_005015.
CC /FtId=VSP_005016.
CC /FtId=VSP_005017.
CC /FtId=VSP_005018.
CC /FtId=VSP_005019.
CC /FtId=VSP_005020.
CC /FtId=VSP_005021.
CC /FtId=VSP_005022.
CC /FtId=VSP_005023.
CC /FtId=VSP_005024.
CC /FtId=VSP_005025.
CC /FtId=VSP_005026.
CC /FtId=VSP_005027.
CC /FtId=VSP_005028.
CC /FtId=VSP_005029.
CC /FtId=VSP_005030.
CC /FtId=VSP_005031.
CC /FtId=VSP_005032.
CC /FtId=VSP_005033.
CC /FtId=VSP_005034.
CC /FtId=VSP_005035.
CC /FtId=VSP_005036.
CC /FtId=VSP_005037.
CC /FtId=VSP_005038.
CC /FtId=VSP_005039.
CC /FtId=VSP_005040.
CC /FtId=VSP_005041.
CC /FtId=VSP_005042.
CC /FtId=VSP_005043.
CC /FtId=VSP_005044.
CC /FtId=VSP_005045.
CC /FtId=VSP_005046.
CC /FtId=VSP_005047.
CC /FtId=VSP_005048.
CC /FtId=VSP_005049.
CC /FtId=VSP_005050.
CC /FtId=VSP_005051.
CC /FtId=VSP_005052.
CC /FtId=VSP_005053.
CC /FtId=VSP_005054.
CC /FtId=VSP_005055.
CC /FtId=VSP_005056.
CC /FtId=VSP_005057.
CC /FtId=VSP_005058.
CC /FtId=VSP_005059.
CC /FtId=VSP_005060.
CC /FtId=VSP_005061.
CC /FtId=VSP_005062.
CC /FtId=VSP_005063.
CC /FtId=VSP_005064.
CC /FtId=VSP_005065.
CC /FtId=VSP_005066.
CC /FtId=VSP_005067.
CC /FtId=VSP_005068.
CC /FtId=VSP_005069.
CC /FtId=VSP_005070.
CC /FtId=VSP_005071.
CC /FtId=VSP_005072.
CC /FtId=VSP_005073.
CC /FtId=VSP_005074.
CC /FtId=VSP_005075.
CC /FtId=VSP_005076.
CC /FtId=VSP_005077.
CC /FtId=VSP_005078.
CC /FtId=VSP_005079.
CC /FtId=VSP_005080.
CC /FtId=VSP_005081.
CC /FtId=VSP_005082.
CC /FtId=VSP_005083.
CC /FtId=VSP_005084.
CC /FtId=VSP_005085.
CC /FtId=VSP_005086.
CC /FtId=VSP_005087.
CC /FtId=VSP_005088.
CC /FtId=VSP_005089.
CC /FtId=VSP_005090.
CC /FtId=VSP_005091.
CC /FtId=VSP_005092.
CC /FtId=VSP_005093.
CC /FtId=VSP_005094.
CC /FtId=VSP_005095.
CC /FtId=VSP_005096.
CC /FtId=VSP_005097.
CC /FtId=VSP_005098.
CC /FtId=VSP_005099.
CC /FtId=VSP_005100.
CC /FtId=V
```

FT VARSPLIC 375 738 Missing (in isoform 2).  
FT CONFLICT 46 46 /FTID=VSP\_004532.  
FT SEQUENCE 738 AA; 82177 MW; 168E8757143BF1F6 CRC64;  
SQ SEQUENCE 738 AA; 82177 MW; 168E8757143BF1F6 CRC64;  
Query Match 3.7%; Score 98.5; DB 1; Length 738;  
Best Local Similarity 21.4%; Pred. No. 2.5; 99; Indels 69; Gaps 10;  
Matches 55; Conservative 34; Mismatches 99; Indels 69; Gaps 10;  
QY 274 IFPAATFOCKN-DIELTPKGYEATFNWENYLEKTSKAAPSRFLNMDCPNHFQKVMKL 332  
DB 441 VFKKTENSENLVDLT---YDIQSFTDVTYRQAINSKM-----FELDMKI 482  
QY 333 EAVDLMPELRLICVATVRVVRHLISIF-----DGM-----DSEVDQWVDESPP 377  
DB 483 AAMHVKKQL-----HQLLPVHLQKHKHSTEGVKLALNDSLDLSMDSDNSM 532  
QY 378 IYPVGMCELTGYQLQ-----PPVAAEPATPLKAKEAT-----KKKKQFGKRRK 421  
DB 533 SVPSPTSAMKTSPLNSSGSGRNSPAPAVTAASVTISQASEVSVPOANSESPGPGPSE 592  
QY 422 RIPTKTRPLRQSGKKLLEDDPGARKISSEVPGEIIIVRVKEHLVY---ASPDKAS 478  
DB 593 SIPQTATQPAISPPPTKTRVSRVSTSLVNPSPRPSGNTATKVPNPVGVKRTSSPNKEE 652  
QY 479 SPELPVSVENIKQETDD 495  
DB 653 SP-----KTKTEDE 663  
RESULT 10  
Z198 HUMAN STANDARD; PRT: 1377 AA.  
AC Q9UBW7; O434212; O43434; O60898; Q9H538; Q9UEU2;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE Zinc finger protein 198 (fused in myeloproliferative disorders  
protein) (Rearranged in atypical myeloproliferative disorder protein).  
GN ZNF198 OR FIM OR RAMP.  
OS Homo sapiens (Human)  
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98245146; PubMed=9576949;  
RA Popovici C., Adiaide J., Ollendorff V., Chaffanet M., Guasch G.,  
RA Jacrot M., Leroux D., Birnbaum D., Pebusque M.-J.;  
RT "Fibroblast growth factor receptor 1 is fused to FIM in stem-cell  
RT myeloproliferative disorder with t(8;13)(p12;q12).";  
RL Proc. Natl. Acad. Sci. U.S.A. 95:5712-5717(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98384233; PubMed=9715603;  
RA Reiter A., Schol J., Kulkarni S., Chase A., Macdonald D.H.C.,  
RA Aguilar R.C.T., Goncalves C., Hernandez J.M., Jennings B.A.,  
RA Goldman J.M., Cross N.C.P.;  
RT "Consistent fusion of ZNF198 to the fibroblast growth factor receptor-  
RT 1 in the t(8;13)(p11;q12) myeloproliferative syndrome.";  
RL Blood 92:1735-1742(1998).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99107818; PubMed=9889006;  
RA Kulkarni S., Reiter A.J., Smedley D., Goldman J.M., Cross N.C.P.;  
RT "The genomic structure of ZNF198 and location of breakpoints in the  
RT t(8;13) myeloproliferative syndrome.";  
RL Genomics 55:118-121(1999).  
RN [4]  
RP SEQUENCE FROM N.A.  
RA Blakey S., Wall M.;  
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
RN [5]

RP SEQUENCE OF 152-1377 FROM N.A.  
RX MEDLINE=98361795; PubMed=9694738;  
RA Still I.H., Cowell J.K.;  
RT "The t(8;13) atypical myeloproliferative disorder: further analysis of  
RT the ZNF198 gene and lack of evidence for multiple genes disrupted on  
RT chromosome 13.";  
RL Blood 92:1456-1458(1998).  
RN [6]  
RP SEQUENCE OF 170-1020 FROM N.A.  
RX MEDLINE=98167848; PubMed=9499416;  
RA Smedley D., Hamoudi R., Clark J., Warren W., Abdul-Rauf M., Somers G.,  
RA Venter D., Fagan K., Cooper C., Shipley J.;  
RT "The t(8;13)(p11;q12) rearrangement associated with an atypical  
RT myeloproliferative disorder fuses the fibroblast growth factor  
RT receptor 1 gene to a novel gene RAMP.";  
RL Hum. Mol. Genet. 7:637-642(1998).  
RN [7]  
RP SEQUENCE OF 621-1377 FROM N.A.  
RX MEDLINE=98085877; PubMed=9425908;  
RA Xiao S., Nalabolu S.R., Aster J.C., Ma J., Abruzzo L., Jaffe E.S.,  
RA Stone R., Weissman S.M., Hudson T.J., Fletcher J.A.;  
RT "FGFR1 is fused with a novel zinc-finger gene, ZNF198, in the t(8;13)  
RT leukaemia/lymphoma syndrome.";  
RL Nat. Genet. 18:84-87(1998).  
CC -!- FUNCTION: May function as a transcription factor.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).  
CC -!- DISEASE: Involved in a t(8;13)(p12;q12) chromosomal translocation  
CC which involves FGFR1 AND ZNF198. The resulting transcript is a  
CC possible candidate for stem cell leukemia lymphoma syndrome/SCLL.  
CC -!- SIMILARITY: Contains 5 MYN-type zinc fingers.  
CC -!- CAUTION: Ref.5 sequence differs from that shown due to a  
CC frameshift in position 330.  
CC -!- CAUTION: Ref.6 sequence differs from that shown due to frameshifts  
CC in positions 330, 966, 1009 and 1017.  
CC -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;  
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/ZNF198ID114.html".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; Y13472; CAA73875.1; -  
CC EMBL; AJ224901; CAA12204.1; -  
CC EMBL; AJ007676; CAA07604.1; -  
CC EMBL; AJ007677; CAA07604.1; JOINED.  
CC EMBL; AJ007678; CAA07604.1; JOINED.  
CC EMBL; AJ007679; CAA07604.1; JOINED.  
CC EMBL; AJ007680; CAA07604.1; JOINED.  
CC EMBL; AJ007681; CAA07604.1; JOINED.  
CC EMBL; AJ007682; CAA07604.1; JOINED.  
CC EMBL; AJ007683; CAA07604.1; JOINED.  
CC EMBL; AJ007684; CAA07604.1; JOINED.  
CC EMBL; AJ007685; CAA07604.1; JOINED.  
CC EMBL; AJ007686; CAA07604.1; JOINED.  
CC EMBL; AJ007687; CAA07604.1; JOINED.  
CC EMBL; AJ007688; CAA07604.1; JOINED.  
CC EMBL; AJ007689; CAA07604.1; JOINED.  
CC EMBL; AJ007690; CAA07604.1; JOINED.  
CC EMBL; AJ007691; CAA07604.1; JOINED.  
CC EMBL; AJ007692; CAA07604.1; JOINED.  
CC EMBL; AJ007693; CAA07604.1; JOINED.  
CC EMBL; AJ007694; CAA07604.1; JOINED.  
CC EMBL; AJ007695; CAA07604.1; JOINED.  
CC EMBL; AJ007696; CAA07604.1; JOINED.  
CC EMBL; AL137119; CAC42467.1; -  
CC EMBL; AL138688; CAC16956.1; -  
CC EMBL; AF060181; AAC23591.1; ALT\_FRAME.  
CC EMBL; AF035374; AAB88464.1; ALT\_FRAME.  
CC EMBL; AF012126; AAC01561.1; -



```
DR PIR; T45119; T45119.
DR Genew; HGNC:12989; ZNF198.
DR MIM; 602221; -.
DR GO; GO:0008270; F-ine ion binding activity; NAS.
KW Transcription regulation; Nuclear protein; Chromosomal translocation;
Repeat; Zinc-finger.
FT ZN_FING 331 409 MYM-TYPE 1 (POTENTIAL).
FT ZN_FING 425 502 MYM-TYPE 2 (POTENTIAL).
FT ZN_FING 536 616 MYM-TYPE 3 (POTENTIAL).
FT ZN_FING 639 712 MYM-TYPE 4 (POTENTIAL).
FT ZN_FING 724 799 MYM-TYPE 5 (POTENTIAL).
FT SITE 913 914 BREAKPOINT FOR TRANSLATION TO FORM
ZNF198-FGFR1.
FT CONFLICT 388 416 VAQDSSEFQECFSTCLSYEDKNPT -> GWLKWIQV
SPSNSVTSCLSLMKNNRIIL (IN REF. 1).
FT CONFLICT 657 659 NKV -> ASL (IN REF. 4).
FT CONFLICT 736 736 K -> G (IN REF. 1).
FT CONFLICT 766 767 MISSING (IN REF. 4).
FT CONFLICT 967 967 MISSING (IN REF. 6).
FT CONFLICT 1009 1010 DF -> IS (IN REF. 6).
FT CONFLICT 1016 1016 MISSING (IN REF. 6).
SQ SEQUENCE 1377 AA; 154910 MW; 2652D4C766492FF9 CRC64;

Query Match 3.6%; Score 96.5; DB 1; Length 1377;
Best Local Similarity 20.2%; Pred. No. 8.4;
Matches 75; Conservative 46; Mismatches 147; Indels 103; Gaps 17;

QY 157 WCHWSPLIHPVGSRRVGHGIRMSRRSDMAHHTPRKIYCDAPVYLFKKVRAVTEGG 216
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 323 NHGKVGKLEAVDLMEPLICVATKRVVHRLLSIHFDG-WSEYDQWDCSPDIYVP 391
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 845 -----MKNKAV-LCKPLTKATYCKPHQTKSCQTDTRTEYVP-VPIPVVYIPV 895
QY 382 GWCELTGQLOPPVAAEATPLKAKATKTKKKQFGKKRKRIPPTKRLRQSGKKPLLE 441
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 896 ---PMHYSQNIPTVTPVFPV-----VPVFLPAPLDSSEKIPAAI 934
QY 442 DDPQARKISSEPVGEIIV--RVKEE-----HLDVASPD--KASSPELP 483
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 935 EELKS--KVSSDALDTELLTMDMSDEGKTEETNTINSVITDIIGSDLLKNSDPETQ 992
QY 484 VSVENIKOETD 494
QY 993 SSMPDVPYEPD 1003

RESULT 11
ENP1 TORCA STANDARD; PRT; 721 AA.
ID ENP1 TORCA STANDARD; PRT; 721 AA.
AC P14400;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Electromotor neuron-associated protein 1 (Fragment).
OS Torpedo californica (Pacific electric ray).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squalia; Hymenoptera; Pristionotidae; Batoidae;
OC Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
OX NCBI_TaxID=7787;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90091742; PubMed=2480872;
RA Ngse J.K., Scheller R.H.;
```

```
RT "Isolation and characterization of two homologous cDNA clones from
RT Torpedo electromotor neurons.";
RL DNA 8:555-561(1989).
CC -|- SURCELLULAR LOCATION: ASSOCIATED WITH MEMBRANES OF INTRACELLULAR
CC ORGANELLES.
CC -|- MISCELLANEOUS: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE
CC SEQUENCE KKE, REPEATED BUT NOT AT FIXED INTERVALS.
CC -|- SIMILARITY: TO ELECTROMOTOR NEURON-ASSOCIATED PROTEIN 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M30270; AAA49279.1; -.
DR PIR; A33319; A33319.
DR Neurone; Repeat.
KW NON_TER 1 719
FT DOMAIN 541 715 11 X 3 AA APPROXIMATE REPEATS.
FT REPEAT 563 565 1.
FT REPEAT 566 568 2.
FT REPEAT 666 668 3.
FT REPEAT 673 675 4.
FT REPEAT 681 683 5.
FT REPEAT 689 691 6.
FT REPEAT 694 696 7.
FT REPEAT 700 702 8.
FT REPEAT 705 707 9.
FT REPEAT 709 711 10.
FT REPEAT 713 715 11.
FT NON_TER 721 721
SQ SEQUENCE 721 AA; 81241 MW; B3E69A9F309CE00 CRC64;

Query Match 3.6%; Score 95.5; DB 1; Length 721;
Best Local Similarity 22.2%; Pred. No. 4.3;
Matches 74; Conservative 45; Mismatches 112; Indels 103; Gaps 18;

QY 232 LGNICV-----ATCKVL-----LDGLMICVDGSPSTGLDWFVCYHSSH 272
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 393 LNKLCVKPEPLFRVGNVIDPVLVFKMGVGRLEMYILNSVKSKE---LQFFMOHWSN 449
QY 273 AIFPATFOQNDIELTPPKGYEAO--TFNWNYLETKS-----KAAPSR-----LFFNMD 321
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 450 -----NKATGILLPDKDAEISF---PYLTSSILVHHPANPSEKIVRALPGNA 498
QY 322 PNHGKVGVM-KLEAVDLMEPLI-----CVATKRVVHRLLSIHFDGWDSEYDQWV 371
DQCHLL-RFYCQNEPNTTKGPENLHYDQGCQTSTRTKMTGSAPPSPTPNKE-- 844
QY 499 POYNILDGLKHLDFLKHFPVVTQKELIASAAPTQK-----AKLKQWS 544
QY 372 D-----CESPDYPCWCELGTGLOPPVAAEATPLKAKATKTKKKQFGKKRKR----- 423
QY 545 DSKESLKSNSRPSVSGKGVKQVKEETPELTTP-TAVSHQEAENVKPKVKEKEKPVVKE 603
QY 424 -----PPTKTRPLRQSGKKPLLEDDPQARKISSE-----PVPGEIIVRV----- 464
QY 604 RPTRELQSKPEEKDAKADAAQAELEKMKQCKELKSESKPKPLKEKIVKEKAKKPE 663
QY 465 --KEHLDV-----ASPDKASSPELPVSVENIKQE 492
QY 664 EKKGEKDYKESAKPDRKEKAV-IXKEKVKKE 696

RESULT 12
CBP5 CANAL STANDARD; PRT; 479 AA.
ID CBP5 CANAL STANDARD; PRT; 479 AA.
AC O43101;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```

```
DE Centromere/microtubule binding protein CBP5 (Centromere-binding factor
DE 5) (Nucleolar protein CBP5).
GN CBP5.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Jiang W., Clifford J., Koltin Y.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS IN VITRO TO CENTROMERES AND MICROTUBULES. IT IS A
CC CENTROMERE DNA-CBP3-BINDING FACTOR AND IS INVOLVED IN MITOTIC
CC CHROMOSOME SEGREGATION. IT IS ESSENTIAL FOR CELL GROWTH. MAY BE IN
CC SOME WAY ASSOCIATED WITH THE CBP3 110 kDa SUBUNIT (CBP3A) (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TRUE FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -!- SIMILARITY: Contains 1 PUA domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U59149; AB894297.1; -.
DR InterPro; IPR004802; Cbf5.
DR InterPro; IPR002478; PUA.
DR InterPro; IPR002501; TruB N.
DR InterPro; IPR004521; Unchar_dom_2.
DR Pfam; PF01472; PUA; 1.
DR Pfam; PF01509; TruB N; 1.
DR SMART; SM00359; PUA; 1.
DR TIGRFAMs; TIGR00425; CBF5; 1.
DR TIGRFAMs; TIGR00451; unchar_dom_2; 1.
DR PROSITE; PS00890; PUA; 1.
DR Microtubules; Centromere; Repeat; Nuclear protein; DNA-binding.
FT DOMAIN 267 342
FT PUA.
SQ SEQUENCE 479 AA; 54321 MW; 3BAF5104E12C9EB6 CRC64;

Query Match 3.4%; Score 92.5; DB 1; Length 479;
Best Local Similarity 20.9%; Pred. No. 4.4;
Matches 103; Conservative 61; Mismatches 155; Indels 173; Gaps 27;

QY 42 NDASHDFWNLGTVDPVHPFGICAINSKILVPPRTIIRAKFTDWKGYLMKLVGSRITLPDVF 101
DB 70 NPSHEV-----VAVI---KRLRVEKTHS-----GTLDPKVTGCLIVCIDR 109
QY 102 HIKVSMKYPFRQGMRLVV-----DKSQVSRTRMAYVDTVIGG----- 141
DB 110 ATRLVKS-----QQGAGKEYVCIVRLHEQLKDKDELNR-----ALENLGALFORPLISA 160
QY 142 -----RLRLIYEDGSDDD-----FWC-----HWSPLIH-----PVGWS-----RRV 174
DB 161 VKQLRVRTVYDVKLIIEFNKRGVLGVFASCEAGTYMRTLCVHLGMLLVGGVGHMQLRVL 220
QY 175 GHGFKMSEKSDMAHPTFRKIYCDAPVLYFKKRVAVYTGEGWFEGMKLEATDPLNLGN 234
DB 221 RSG-AMSESDNIVTLHVL-----DA-QYVDYDNR-----DESYLEKLIQPL----- 260
QY 235 ICVATVCKVLLDGYLMTCVDPGSGSTDGLMFCFHASHAIFATFCQKNDIELTTPKGYE 294
DB 261 -----ESLLVGYKRVVW-----KDSAVNSVCYGAKE--LMIPGLLYBERGIEL-----YD 302
QY 295 AQTFNWNYLEKTKSKAAPSRLENMDCPNHGFVKGM-KLEAVLMEPRLICVATVCKR--- 350
DB 303 -----EVLMTTKGEA-----IAIGQMTVDLQSCDHGIYAKVRCRM 342
QY 351 -----VVHRLSIHFDGWDSEYDQWDCSDPDYIPVGHGCELTGYQLQPP--- 394
```

```
DB 343 EBDTPYRWGLPIAQKKKKQKADKLDKYGR-VNENTPENWKKDYKDL-DEQPAPPIPE 400
QY 395 ---VAAPATPLKA-----KEATKKKKKQFGKKRIPPTKTRPRRQSGKK 437
DB 401 SKLVAPEQLPKKSLIBEVEDKSKGKKKKKKKKKKKKKKKKKKKKKKKKKK 460
QY 438 PLEDDPQGARK 449
DB 461 RKAEDDSKSEK 472

RESULT 13
AT12 HUMAN STANDARD; PRT; 1593 AA.
ID P58397.
AC 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ADAMTS-12 precursor (EC 3.4.24.-) (A disintegrin and metalloproteinase
DE with thrombospondin motifs 12) (ADAM-TS 12) (ADAM-TS12).
GN ADAMTS12.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Petal lung;
RX MEDLINE=21264577; PubMed=11279086;
RA Cal S., Argueelles J.M., Fernandez P.L., Lopez-Otin C.;
RT "Identification, characterization, and intracellular processing of
RT ADAM-TS12, a novel human disintegrin with a complex structural
RT organization involving multiple thrombospondin-1 repeats.";
RL J. Biol. Chem. 276:17932-17940(2001).
CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Secreted. Associated with the extracellular
CC matrix (By similarity).
CC -!- TISSUE SPECIFICITY: Expressed exclusively in fetal lung. Is widely
CC expressed in gastric carcinomas and in cancer cells of diverse
CC origin.
CC -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX (BY
CC SIMILARITY).
CC -!- PTM: THE PRECURSOR IS CLEAVED BY A PURIN ENDOPEPTIDASE.
CC -!- PTM: IS SUBJECT TO AN INTRACELLULAR MATURATION PROCESS LEADING
CC TO A FRAGMENT CONTAINING THE N-TERMINAL REGION INCLUDING THE
CC METALLOPROTEINASE, DISINTEGRIN-LIKE, CYS-RICH AND TS-1 DOMAINS AND
CC THE C-TERMINAL FRAGMENT CONTAINING THE SPACER 2 AND THE FOUR TS-1
CC DOMAINS.
CC -!- SIMILARITY: Belongs to peptidase family M12B.
CC -!- SIMILARITY: Contains 1 disintegrin-like domain.
CC -!- SIMILARITY: Contains 1 PLAC domain.
CC -!- SIMILARITY: Contains 8 TSP type-1 domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ250735; CAC20419.1; -.
DR Genbank; HGNC:14605; ADAMTS12.
DR MIM; 606184; -.
DR InterPro; IPR001762; Disintegrin.
DR InterPro; IPR002870; Pep M12B propep.
DR InterPro; IPR001590; Reprolysin.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR006025; Zn_Mtpeptdse.
DR Pfam; PF01562; Pep M12B propep; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF00090; tsp_1; 6.
```

DR SMART; SM00209; TSP1; 8.  
 DR PROSITE; PS00215; ADAM MEPRO; 1.  
 DR PROSITE; PS00546; CYSTEINE SWITCH; FALSE NEG.  
 DR PROSITE; PS00427; DISINTEGRIN 1; FALSE NEG.  
 DR PROSITE; PS00214; DISINTEGRIN 2; FALSE NEG.  
 DR PROSITE; PS00092; TSP1; 6.  
 DR PROSITE; PS00142; ZINC PROTEASE; 1.  
 KW Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;  
 KW Repeat; Extracellular matrix.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT PROPEP 26 240 BY SIMILARITY.  
 FT CHAIN 241 1593 ADAMTS-12.  
 FT DOMAIN 241 464 METALLOPROTEASE.  
 FT DOMAIN 465 544 DISINTEGRIN-LIKE.  
 FT DOMAIN 542 597 TSP TYPE-1 1.  
 FT DOMAIN 597 700 CYS-RICH.  
 FT DOMAIN 701 826 SPACER 1.  
 FT DOMAIN 823 882 TSP TYPE-1 2.  
 FT DOMAIN 886 942 TSP TYPE-1 3.  
 FT DOMAIN 943 996 TSP TYPE-1 4.  
 FT DOMAIN 996 1315 SPACER 2.  
 FT DOMAIN 1312 1365 TSP TYPE-1 5.  
 FT DOMAIN 1367 1421 TSP TYPE-1 6.  
 FT DOMAIN 1422 1470 TSP TYPE-1 7.  
 FT DOMAIN 1471 1531 TSP TYPE-1 8.  
 FT DOMAIN 302 305 POLY-GLU.  
 FT DOMAIN 1538 1570 PLAC.  
 FT METAL 208 208 CYSTEINE SWITCH (POTENTIAL).  
 FT SITE 392 392 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT ACT SITE 393 393 BY SIMILARITY.  
 FT METAL 396 396 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT METAL 402 402 ZINC (CATALYTIC) (BY SIMILARITY).  
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 215 215 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1104 1104 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1275 1275 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1300 1300 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1320 1320 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1371 1371 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1378 1378 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 1503 1503 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SQ SEQUENCE 1593 AA; 177545 MW; 07P9P48E63B83A3 CRC64;  
 Query Match 3.4%; Score 92.5; DB 1; Length 1593;  
 Best Local Similarity 18.4%; Pred. No. 22;  
 Matches 93; Conservative 59; Mismatches 166; Indels 187; Gaps 26;  
 QY 9 LNSDAVLPSRYVWTASVIQTAGYRVLLRYSGFENDAS----HDFWCNLG-TVDVHPIGWC 63  
 DB 378 INDSGLPLAF---TIAHELHSGFQHDGKENDCEPVRHRYIMSRLOQYDPTPLTWS 433  
 QY 64 AINSKILVPPRTTHAKFTD--WKGYLMKRLVGSRTLPDPFHIDQVSMKYPYPRQGMRLSV 121  
 DB 434 KCSEYI-----TRFLDRGW-GFCLDDI-----PKKGLKSKV 465  
 QY 122 V-----DKSQVSRTRMAVMTVIGRLLRYEDGSDDDDFW 157  
 DB 466 IAPGVYDVHQQLOLYGNPNATFCQEVENYQTLWCSSVKGFCSRKLDAADAGTQCCKW 525  
 QY 158 C-----HMSPLTHPVGWRRVGHGKIMSER-----183  
 DB 526 CMACKCITVGKPPSIPGGWGRSPWSH---CSTCGAGVQSAERLCLNNPPEPKGGKYCT 582  
 QY 184 -----RSDMAHPTPRKTYC---DAVPY-----LFFKVR--AVYTE- 214  
 DB 583 GERKRYRLCNVHPCKRSEA---PTFRMQCSEFTVPYKNELVHWFPIFNPAHPCELYCRP 639  
 QY 215 -GGWFEESGKLEADPL-----NLGNICVATVCKVLLDGYLMICVDGGPSTDGLDWFY 267

640 IDGQFSEKMLDAVIDGTPCFEGGNSRVNCINGICKWVGCDYE---IDSNATEDRCGCVCLG 696  
 268 HASSHAIFPATFOCKN-----DIELTPPKGYEAQTENWE---NYLEKTKSKAAPSRLENNM 319  
 697 DGSSCQIVRKFKQKSGSGYVDGLPKGARDIRVMEIEGAGNFLAIRSED--PEKY-- 752  
 320 DCPNHGFKV---GMKLEAVDL-----MEPLRICVATVVRVHRL----- 356  
 753 --LNGGFIIOWNGNYKLAGTVFQYDRKGDLKLMATGPTNESVWIIQLLFQVTPNGIKYEY 810  
 357 SIHEDGWDG-----EYDQWDC 373  
 811 TIQDGLDNDVEQMYFWQYGHWTCC 835

RESULT 14  
 HL MAIZE  
 ID HL MAIZE  
 AC P23444; STANDARD; PRT; 245 AA.  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Histone H1.  
 OS Zea mays (Maize).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACAD clade; Panicoideae; Andropogoneae; Zea.  
 OX NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. W22;  
 RX MEDLINE=91227140; PubMed=1709276;  
 RA Raxafmahatratra P., Chaubet N., Philipps G., Gigot C.;  
 RT "Nucleotide sequence and expression of a maize H1 histone cDNA."  
 RL Nucleic Acids Res. 19:1491-1496(1991).  
 CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF  
 CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X57077; CAA40362.1; -.  
 CC PIR; S26826; S26826.  
 CC MaizeDB; 25540; -.  
 CC InterPro; IPR005818; Histone H1/H5.  
 CC InterPro; IPR005819; Histone H5.  
 CC InterPro; IPR003216; Linkerhist.N.  
 CC Pfam; PF00538; linker\_histone; 1.  
 CC PRINTS; PR00624; HISTONEH5.  
 CC ProDom; PD000373; Linkerhist\_N; 1.  
 CC SMART; SM00526; H15; 1.  
 CC Chromosomal protein; Nuclear protein; DNA-binding.  
 FT INIT MET 0 0  
 FT DOMAIN 1 30 ASP/GLU-RICH (ACIDIC).  
 FT DOMAIN 31 45 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 46 121 GLOBULAR.  
 SQ SEQUENCE 245 AA; 25217 MW; D1FD73DAA7A6778 CRC64;  
 Query Match 3.4%; Score 91.5; DB 1; Length 245;  
 Best Local Similarity 33.3%; Pred. No. 2.2;  
 Matches 34; Conservative 12; Mismatches 37; Indels 19; Gaps 5;  
 QY 392 QPVAEEP--ATPLKAEATKKK-----KKQFGKKRGRIPPTKTRPLRQGSK-KPL 439  
 DB 139 KPAAAKPKAKTPAKAKPATPKPAKPAKVAWKP-KTPAKPKAKPAKPAKTAGAKPPL 198



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 47 Seconds  
(without alignments)  
1671.695 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKEVLNSDAVLPSRVY.....KASSPELPVSVNIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Genesepc\_19Jun03:\*  
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2683	100.0	495	22 AAM39576	Human polypeptide
2	2683	100.0	495	22 AAB60488	Human cell cycle a
3	2436	90.8	448	22 AAM41362	Human polypeptide
4	2185	81.4	617	23 AEB05680	Human transmembran
5	1688	62.9	329	23 AEB89448	Human polypeptide
6	1410	52.6	631	21 AAY79183	Haematopoietic ste
7	1395	52.0	410	24 ABU03557	Angiogenesis-assoc
8	1387	51.7	410	23 ABB78695	Human hsp BAA90919
9	1111	41.4	891	22 ABB70683	Drosophila melanog

10	1111	41.4	1243	22 ABB65759	Drosophila melanog
11	620.5	23.1	866	23 ABB97494	Novel human protei
12	611	22.8	894	23 ABB69567	Human polypeptide
13	587	21.9	205	22 ABB14398	Novel human diagn
14	510	19.0	538	24 ABB11746	Human MDT polypep
15	510	19.0	538	24 ABB11833	Human MDT polypep
16	507	18.9	826	20 AAY01072	Rat 1(3)mbt protei
17	505	18.8	826	20 AAY01071	Mouse 1(3)mbt prot
18	501.5	18.7	738	20 AAY01070	Human 1(3)mbt prot
19	501.5	18.7	738	20 AAY01069	Human 1(3)mbt prot
20	494.5	18.4	498	23 ABB78650	Caenorhabditis ele
21	493.5	18.4	498	23 ABB78650	C. elegans mutant
22	486.5	18.1	498	23 ABB78699	C. elegans mutant
23	471	17.6	713	22 AAM00891	Human bone marrow
24	471	17.6	713	22 AAM00892	Human bone marrow
25	417	15.5	77	22 ABB53185	Human liver peptid
26	417	15.5	77	22 ABB38351	Peptide #5857 enco
27	417	15.5	77	22 ABB23528	Peptide #5827 enco
28	417	15.5	77	22 AAM58963	Protein #5527 enco
29	417	15.5	77	22 AAM71490	Human brain expres
30	417	15.5	77	22 AAM19142	Human bone marrow
31	417	15.5	77	22 AAM19142	Peptide #5576 enco
32	417	15.5	77	22 AAM19142	Peptide #5827 enco
33	402	15.0	1477	22 ABB61960	Human peptide enco
34	373.5	13.9	217	23 ABB41407	Drosophila melanog
35	364.5	13.6	591	18 AAW37399	Human ovarian anti
36	364.5	13.6	591	18 AAW37395	Human sex comb on
37	364	13.6	623	22 ABB03903	Human sex comb on
38	362.5	13.5	269	24 ABB11777	Novel human diagn
39	359	13.4	246	24 ABB11603	Human MDT polypep
40	357.5	13.3	591	20 AAY18289	Human MDT polypep
41	357.5	13.3	620	18 AAW37397	Human Scm protein
42	357.5	13.3	620	18 AAW37393	Human sex comb on
43	357.5	13.3	620	18 AAW37393	Human sex comb on
44	355.5	13.3	648	22 AAB94447	Human Scm protein
45	353.5	13.2	577	18 AAW37398	Human protein sequ
					Human sex comb on

## ALIGNMENTS

RESULT 1  
AAM39576 standard; Protein, 495 AA.  
ID AAM39576 standard; Protein, 495 AA.  
XX AC AAM39576;  
XX DT 22-OCT-2001 (first entry)  
XX DE Human polypeptide SEQ ID NO 2721.  
XX KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX KW peripheral nervous system, neuropathy; central nervous system; CNS;  
XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX KW myoelectric lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX KW leukaemia.  
XX OS Homo sapiens.  
XX PN WO200153312-A1.  
XX PD 26-JUL-2001.  
XX PF 26-DEC-2000; 2000WO-US34263.  
XX PR 21-JAN-2000; 2000US-0488725.  
XX PR 25-APR-2000; 2000US-052317.  
XX PR 09-JUL-2000; 2000US-0598042.  
XX PR 13-JUL-2000; 2000US-0620312.  
XX PR 03-AUG-2000; 2000US-0653450.  
XX PR 14-SEP-2000; 2000US-0662191.  
XX PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.  
 PA (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 FI Zhao QA, Zhou F, Goodrich R, Drmarac RT;  
 XX WPI; 2001-442253/47.  
 DR N-PSDB; AAI58732.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 4; SEQ ID NO 2721; 10078pp; English.  
 CC The invention relates to human nucleic acids (AAI57798-AAI61359) and  
 CC the encoded polypeptides (AAI38642-AAI42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 495 AA;  
 Query Match 100.0%; Score 2683; DB 22; Length 495;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-269;  
 Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MKGMKVELNSDAVLPBRYVWIASVITAGYRVLLRYEGFENDASHDFWNCNLGTVDVHP 60  
 Db 1 MKGMKVELNSDAVLPBRYVWIASVITAGYRVLLRYEGFENDASHDFWNCNLGTVDVHP 60  
 Qy 61 GWCAINSKILVPPRTIHAKTIDMKYLMKRLVGSRTLPVDFHFKWVESKKYPPQGMRL 120  
 Db 61 GWCAINSKILVPPRTIHAKTIDMKYLMKRLVGSRTLPVDFHFKWVESKKYPPQGMRL 120  
 Qy 121 VVDKSOVSRTMAVVDTVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 180  
 Db 121 VVDKSOVSRTMAVVDTVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 180  
 Qy 181 SERSDMAHPTFKIYCDVAVPLFKKRVAVYEGGFEGMKLEAIDPLNLGNICVATV 240  
 Db 181 SERSDMAHPTFKIYCDVAVPLFKKRVAVYEGGFEGMKLEAIDPLNLGNICVATV 240  
 Qy 241 CKVLIDGYLMICVDGSGSTDGLDWCFCVHSSHAIFPATFCQKNDIELTPKGYEATFNW 300  
 Db 241 CKVLIDGYLMICVDGSGSTDGLDWCFCVHSSHAIFPATFCQKNDIELTPKGYEATFNW 300  
 Qy 301 ENYLEKTSKAAPRLNMDCPNHPGFKVGMKLEAVDLMEPRLLICVATKRVVHRLLSIH 360  
 Db 301 ENYLEKTSKAAPRLNMDCPNHPGFKVGMKLEAVDLMEPRLLICVATKRVVHRLLSIH 360  
 Qy 361 DGNWSEVDQWDCSPDIYPVWCELTGYQLQPPVAEPATPLKAEATKKKKQFGKKR 420  
 Db 361 DGNWSEVDQWDCSPDIYPVWCELTGYQLQPPVAEPATPLKAEATKKKKQFGKKR 420  
 Qy 421 KRIPPTKTRPLRQSKKPLLEDDPQGARKISSEFPVGEIITAVRKGHELDVSPDKASSP 480  
 Db 421 KRIPPTKTRPLRQSKKPLLEDDPQGARKISSEFPVGEIITAVRKGHELDVSPDKASSP 480  
 Qy 481 ELPVSVENIKQETDD 495  
 |||||

Db 481 ELPVSVENIKQETDD 495  
 RESULT 2  
 AAB60488  
 ID AAB60488 standard; Protein; 495 AA.  
 XX  
 AC AAB60488;  
 XX  
 DT 24-APR-2001 (first entry)  
 XX  
 DE Human cell cycle and proliferation protein CCYPR-36, SEQ ID NO:36.  
 KW Cell cycle and proliferation protein; CCYPR; human; agonist;  
 KW antagonist; gene therapy; detection; immune disorder;  
 KW transgenic animal disease model; immune disorder;  
 KW developmental disorder; cell signalling disorder;  
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;  
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;  
 XX menstrual cycle disorder; bacterial infection.  
 OS Homo sapiens.  
 XX WO200107471-A2.  
 PN 01-FEB-2001.  
 PD 21-JUL-2000; 2000WO-US19948.  
 XX 21-JUL-1999; 99US-0145075.  
 PR 08-SEP-1999; 99US-0153129.  
 PR 10-NOV-1999; 99US-0184647.  
 XX (INCY-) INCYTE GENOMICS INC.  
 PA Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
 PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
 XX WPI; 2001-112727/12.  
 DR N-PSDB; AAF59625.  
 XX Human cell cycle and proliferation proteins and polynucleotides are  
 PT used to treat, diagnose and prevent immune, developmental and cell  
 PT signalling disorders and cell proliferative disorders including cancer -  
 XX Claim 1; Page 142-143; 205pp; English.  
 CC Sequences AAB60453-AAB60506 represent 54 human cell cycle and  
 CC proliferation proteins (CCYPR), which are encoded by AAF59590-AAF59643.  
 CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
 CC associated with decreased expression of functional CCYPR, while CCYPR  
 CC antagonists are used to treat diseases or conditions associated with  
 CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
 CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
 CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
 CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
 CC that specifically bind to CCYPR, and in drug screening methods to  
 CC identify compounds that modulate the activity of CCYPR. CCYPR  
 CC nucleotides can be used to generate transgenic animal models of human  
 CC disease, and can be used in gene therapy in target cells with genetic  
 CC abnormalities with respect to the expression of CCYPR for the  
 CC treatment or prevention of a disorder associated with CCYPR.  
 CC Diseases which can be diagnosed, treated and prevented using CCYPR  
 CC proteins, nucleic acids, agonists or antagonists include immune,  
 CC developmental and cell signalling disorders, and cell proliferative  
 CC disorders including cancer. Specific examples of these disorders  
 CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
 CC diabetes mellitus, disorders of the menstrual cycle and infections  
 CC caused by bacteria.  
 XX Sequence 495 AA;  
 SQ  
 Query Match 100.0%; Score 2683; DB 22; Length 495;



Best Local Similarity 100.0%; Pred. No. 1.5e-269;			
Matches 495; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	MKGKVEVLNSDAVLPSRVVYIASVIQTAGYRVLRYEGFENDASHDFWCNLTGTVDVVHPI	60
Db	1	MKGKVEVLNSDAVLPSRVVYIASVIQTAGYRVLRYEGFENDASHDFWCNLTGTVDVVHPI	60
Qy	61	GWCAINSKILVPPRTTHAKTDMKGYLMKLVGSRTLPVDFHIKQVESMKYPPFQGMRL	120
Db	61	GWCAINSKILVPPRTTHAKTDMKGYLMKLVGSRTLPVDFHIKQVESMKYPPFQGMRL	120
Qy	121	VVKSQSVSRTRMAVDVTVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGK	180
Db	121	VVKSQSVSRTRMAVDVTVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGK	180
Qy	181	SRRSDMAHPTFRKIYCDVAVPYLFKKVAVYTEGGWFEGMKLEAIDPLNLGNICVATV	240
Db	181	SRRSDMAHPTFRKIYCDVAVPYLFKKVAVYTEGGWFEGMKLEAIDPLNLGNICVATV	240
Qy	241	CKVLLDGYLMICVDGGSTGDLWFCYHASSHAIFFPATCQKNDIELTPPKGYEATFNW	300
Db	241	CKVLLDGYLMICVDGGSTGDLWFCYHASSHAIFFPATCQKNDIELTPPKGYEATFNW	300
Qy	301	ENYLEKTSKAAPSRLEFMDCPNHFVGMKLEAVDLMEPRLICVATVRRVHRLLSIH	360
Db	301	ENYLEKTSKAAPSRLEFMDCPNHFVGMKLEAVDLMEPRLICVATVRRVHRLLSIH	360
Qy	361	DGWDSEVDQWDCSPDIYPVGMCELTGYQLQPPVAAEPATPKAKEATKKKKQFGK	420
Db	361	DGWDSEVDQWDCSPDIYPVGMCELTGYQLQPPVAAEPATPKAKEATKKKKQFGK	420
Qy	421	KRIPPTKTRPLRQSKKPLEDDPQGARKISSEPVPGELIAVVRKEHLDVASPDKASP	480
Db	421	KRIPPTKTRPLRQSKKPLEDDPQGARKISSEPVPGELIAVVRKEHLDVASPDKASP	480
Qy	481	ELPVSVENIKOETDD 495	
Db	481	ELPVSVENIKOETDD 495	
RESULT 3			
AA41362			
ID AA41362 standard; Protein; 448 AA.			
AC	AA41362;		
XX			
DT	22-OCT-2001 (first entry)		
DE	Human polypeptide SEQ ID NO 6293.		
XX			
KW	Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		
OS	Homo sapiens.		
XX			
EN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
FF	26-DEC-2000; 2000WO-0834263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-052317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		

XX	(HYSE-) HYSEQ INC.				
PA					
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;				
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;				
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;				
XX	WPI; 2001-442253/47.				
DR	N-PSDB; AA160518.				
XX	Novel nucleic acids and polypeptides, useful for treating disorders				
PT	such as central nervous system injuries -				
XX	Example 2; SEQ ID NO 6293; 10078pp; English.				
XX	The invention relates to human nucleic acids (AA157798-AA161369) and				
CC	the encoded polypeptides (AA157798-AA161369) with nontropic,				
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful				
CC	in gene therapy. A composition containing a polypeptide or polynucleotide				
CC	of the invention may be used to treat diseases of the peripheral nervous				
CC	system, such as peripheral nervous injuries, peripheral neuropathy and				
CC	localised neuropathies and central nervous system diseases, such as				
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic				
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the				
CC	utilisation of the activities such as: Immune system suppression,				
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic				
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,				
CC	assays for receptor activity, arthritis and inflammation, leukaemia and				
CC	C.N.S disorders.				
CC	Note: The sequence data for this patent did not form part of the printed				
CC	specification.				
XX	Sequence 448 AA;				
SQ					
	Query Match 90.8%; Score 2436; DB 22; Length 448;				
	Best Local Similarity 99.8%; Pred. No. 6.1e-244;				
	Matches 447; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Qy	48	FWCNGTVDVHPHIGWCAINSKILVPPRTTHAKTDMKGYLMKRLVGSRTLPVDFHIKQVE	107		
Db	1	FWCNGTVDVHPHIGWCAINSKILVPPRTTHAKTDMKGYLMKRLVGSRTLPVDFHIKQVE	60		
Qy	108	SMKYFPRQGMLEVVDSQSVSRTRMAVVDTVIGRLRLLYEDGSDDDFWCHWMSPLIHP	167		
Db	61	SMKYFPRQGMLEVVDSQSVSRTRMAVVDTVIGRLRLLYEDGSDDDFWCHWMSPLIHP	120		
Qy	168	VGWSRRVGHGIMKERSSDMAHPTFRKIYCDVAVPYLFKKVAVYTEGGWFEGMKLEAI	227		
Db	121	VGWSRRVGHGIMKERSSDMAHPTFRKIYCDVAVPYLFKKVAVYTEGGWFEGMKLEAI	180		
Qy	228	DPLNLGNICVATVCKVLLDGYLMICVDGGSTGDLWFCYHASSHAIFFPATCQKNDIEL	287		
Db	181	DPLNLGNICVATVCKVLLDGYLMICVDGGSTGDLWFCYHASSHAIFFPATCQKNDIEL	240		
Qy	288	TPPKGYEATFNWENYLEKTSKAAPSRLEFMDCPNHGFVKGMKLEAVDLMEPRLICVAT	347		
Db	241	TPPKGYEATFNWENYLEKTSKAAPSRLEFMDCPNHGFVKGMKLEAVDLMEPRLICVAT	300		
Qy	348	VKRVRHRLLSIHFGWDSYDQWDCSPDIYPVHGCELTGTQLQPPVAAEPATPKAKE	407		
Db	301	VKRVRHRLLSIHFGWDSYDQWDCSPDIYPVHGCELTGTQLQPPVAAEPATPKAKE	360		
Qy	408	ATKKKKQFGKKRRIIPPTKTRPLRQSKKPLEDDPQGNARKISSEPPVGEIIIAVRKEE	467		
Db	361	ATKKKKQFGKKRRIIPPTKTRPLRQSKKPLEDDPQGNARKISSEPPVGEIIIAVRKEE	420		
Qy	468	HLDVASPDKASSPELPVSVENIKOETDD 495			
Db	421	HLDVASPDKASSPELPVSVENIKOETDD 448			

ID ABB05680 standard; Protein; 617 AA.  
 AC ABB05680;  
 DT 30-APR-2002 (first entry)  
 DE Human transmembrane protein clone amy2\_1114.  
 XX Human; foetal brain; foetal kidney; melanoma; testis; amygdala;  
 XX gene therapy; chromosome 22q13.31-13.33.  
 XX Homo sapiens.  
 OS WO200198454-A2.  
 PN 27-DEC-2001.  
 PD  
 XX 25-APR-2001; 2001WO-1B02050.  
 XX 25-APR-2000; 2000US-199380P.  
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 PA Wiemann S;  
 PI WPI; 2002-055860/07.  
 DR N-PSDB; ABA93718.  
 XX Human cDNA sequences and clones derived from human fetal brain, fetal  
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic  
 PT screening and therapy -  
 XX  
 PS Claim 1; Page 130; 611pp; English.  
 XX The present invention describes assemblies and computer readable media  
 CC comprising novel human cDNA sequences and clones derived from human  
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA  
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the  
 CC present invention which encode the proteins given in ABB05662 to  
 CC ABB05729. The human cDNA sequences and clones can be used in gene  
 CC therapy. The clones may be used in a variety of applications, for  
 CC example they may be used in profiling assays, for providing large arrays  
 CC of human genetic material for implementing large-scale screening  
 CC strategies and for treating diseases via gene therapy procedures.  
 XX  
 SQ Sequence 617 AA;  
 Query Match 81.4%; Score 2185; DB 23; Length 617;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-217;  
 Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MKGMKVEVLNSDAVLPSPRYWIASVITQAGYRVLLRYEGFENDASHDFWCLNLTGVVHPI 60  
 DB 211 MKGMKVEVLNSDAVLPSPRYWIASVITQAGYRVLLRYEGFENDASHDFWCLNLTGVVHPI 270  
 QY 61 GWCAINSKILVPRTHIAKFTDQKYLKRLVGSRTLPVDFHFKMVESMKYFFRQGMRL 120  
 DB 271 GWCAINSKILVPRTHIAKFTDQKYLKRLVGSRTLPVDFHFKMVESMKYFFRQGMRL 330  
 QY 121 VDKSQVSRTRMAVDTVIGGRLLLYEDGSDDDDFWCHWSPPLHPVGSRRVGHGIM 180  
 DB 331 VDKSQVSRTRMAVDTVIGGRLLLYEDGSDDDDFWCHWSPPLHPVGSRRVGHGIM 390  
 QY 181 SERSDMAHPTFRKLYCDAPVPLFKKRAVYTEGGWFEGMKLEAIDPLNLGNTCVATV 240  
 DB 391 SERSDMAHPTFRKLYCDAPVPLFKKRAVYTEGGWFEGMKLEAIDPLNLGNTCVATV 450  
 QY 241 CKVLLDGYLMICVDGGPSTDLGDFWFCYHASSHAIFPATFCQKNDIELTPPKYEAQTNW 300  
 DB 451 CKVLLDGYLMICVDGGPSTDLGDFWFCYHASSHAIFPATFCQKNDIELTPPKYEAQTNW 510  
 QY 301 ENYLEKTSKAAPSRLFNMDCPNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIHF 360

Db 511 ENYLEKTSKAAPSRLFNMDCPNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSIHF 570  
 QY 361 DGWDSYDQWDCESPDYIPVGVWCELTYGLOPPVAA 397  
 Db 571 DGWDSYDQWDCESPDYIPVGVWCELTYGLOPPVAA 507  
 RESULT 5  
 ABB89448  
 ID ABB89448 standard; Protein; 329 AA.  
 XX AC ABB89448;  
 XX 24-MAY-2002 (first entry)  
 XX Human polypeptide SEQ ID NO 1824.  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 XX anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;  
 XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 XX neurological disease; infection; human; secreted protein.  
 OS Homo sapiens.  
 XX WO200190304-A2.  
 XX 29-NOV-2001.  
 XX 18-MAY-2001; 2001WO-US16450.  
 XX 19-MAY-2000; 2000US-205515P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Birse CE, Rosen CA;  
 XX WPI; 2002-122018/16.  
 XX N-PSDB; ABL89857.  
 XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX  
 PS Claim 11; SEQ ID NO 1824; 2081pp + Sequence Listing; English.  
 CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
 CC (b) immune disorders e.g. Addison's disease, diabetes mellitus, Crohn's  
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, ulcerative  
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
 CC and parasitic infections.  
 CC Note: The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 329 AA;  
 Query Match 62.9%; Score 1688; DB 23; Length 329;  
 Best Local Similarity 99.0%; Pred. No. 2.1e-166;  
 Matches 308; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKGMKVEVLNSDAVLPSPRYWIASVITQAGYRVLLRYEGFENDASHDFWCLNLTGVVHPI 60



PR 14-FEB-2001; 2001US-0784356.  
 PR 22-FEB-2001; 2001US-0791390.  
 PR 19-APR-2001; 2001US-285475P.  
 PR 03-AUG-2001; 2001US-310025P.  
 PR 13-NOV-2001; 2001US-350866P.  
 PR 29-NOV-2001; 2001US-334244P.  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX PA  
 XX Murray R, Glynn R, Watson SR, Aziz N;  
 XX WPI; 2003-040691/03.  
 DR N-PSDB; ABX08841.  
 XX  
 PT Detecting angiogenesis-associated transcript in a cell for diagnosing  
 PT and treating cancer by contacting a sample with a polynucleotide that  
 PT exhibits changes in expression level as a function of time in tissue  
 PT undergoing angiogenesis -  
 XX  
 FS Example 2; Page 275; 291pp; English.  
 XX  
 CC The present invention relates to methods and compositions for  
 CC detecting an angiogenesis-associated transcript in a cell in  
 CC a patient. The method involves contacting a biological sample from  
 CC the patient with a polynucleotide that selectively hybridizes to a  
 CC sequence at least 80% identical to any of the angiogenesis-associated  
 CC human polynucleotide sequences given in the specification. These  
 CC angiogenesis-associated polynucleotide sequences comprise genes that  
 CC exhibit changes in expression levels as a function of time in tissue  
 CC undergoing angiogenesis. The method and the polynucleotide sequences  
 CC of the invention are useful for diagnosing and treating angiogenesis  
 CC and angiogenesis-associated diseases e.g. cancer. The polynucleotide  
 CC sequences are also useful in the gene therapy of such disorders. The  
 CC angiogenesis-associated proteins encoded by the polynucleotide  
 CC sequences are useful as a vaccine for therapeutic and prophylactic  
 CC immunisation. ABU03456-ABU03569 represent angiogenesis-associated  
 CC protein sequences.  
 XX  
 XX Sequence 410 AA;  
 SQ  
 Query Match 52.0%; Score 1395; DB 24; Length 410;  
 Best Local Similarity 62.9%; Pred. No. 8.5e-136;  
 Matches 249; Conservative 55; Mismatches 84; Indels 8; Gaps 2;  
 QY 2 KGMKVELNSDAVLPSRVYVIASTVQAGYVLLRYEGFENDASHDFWNCNGTVDVHPIG 61  
 Db 10 ENVRVEVPNTDCLPTKFWIAGIYKLVKLTGAKTLPPDFSKVSESQYFPCMRVEV 69  
 QY 62 WCAINSKILVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESQYFPCMRVEV 121  
 Db 70 WCAASGKPLVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESQYFPCMRVEV 129  
 QY 122 VDKSQSVTRMAVDVTVIGRRLRYEDG-DDDDFWCHMSPLIHPVWGRVGHGKIM 180  
 Db 130 VDKRLHCLTRVAVSVIGRRLRYEEDRTDDFWCHMSPLIHPVWGRVGHGKIM 189  
 QY 181 SRRSDMAHPTFRKTYCDAVYLPFKVRAVYTEGGWFEKGKLEAIDPLNLGICVATV 240  
 Db 190 SDITKKQDGH-----FTTPHLFAKVEVDQSENFKEGKLEAIDPLNLGICVATV 242  
 QY 241 CKVLDGYLMICVDGSPSTDGLDFWCHYHASSHAIPATFOQNDIELPPPKGYEAQTNW 300  
 Db 243 RKVLADGFLMIGDGEADGSDWFCYHATSPSIFPVGFCEINMIELTPPRGYTKLPFKW 302  
 QY 301 ENYLEKTSKAAPSRLENMDCFNHGVKGMKLEAVDLMEPLICVATVYKRVHRLLSHF 360  
 Db 303 FYLRBTSGIAAPVKLVKNDVNHGFRVGMKLEAVDLMEPLICVATVYKRVHRLLSHF 362  
 QY 361 DQWDEYDQWDCESPDIYPVGCWELTGYQLQPPVA 396  
 Db 363 DQWEEYDQWDCESPDIYPVGCWELTGYQLQPPAS 398

RESULT 8  
 ABB78695  
 ID ABB78695 standard; Protein; 410 AA.  
 XX  
 AC ABB78695;  
 XX  
 DT 17-JUL-2002 (first entry)  
 XX  
 DE Human hsp BAA90919 protein sequence SEQ ID NO:68.  
 XX  
 KW Caenorhabditis elegans; lin-8; lin-56; lin-61; tumour suppressor;  
 KW cell proliferation; nematode; cancer; mutant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200194545-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 01-JUN-2001; 2001WO-US17909.  
 XX  
 PR 02-JUN-2000; 2000US-208802P.  
 XX  
 PA (WASI ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 PI Horvitz HR, Davison EM, Lu X;  
 XX WPI; 2002-401590/43.  
 XX  
 PT New Caenorhabditis elegans nucleic acid involved in tumor suppressor  
 PT pathway, encoding LIN-8, LIN-56 or LIN-61 polypeptide, useful for  
 PT diagnosing and treating cell proliferative diseases such as cancer -  
 XX  
 PS Disclosure; Fig 8; 116pp; English.  
 CC  
 CC The present invention describes a substantially pure nematode C. elegans  
 CC (Caenorhabditis elegans) nucleic acid (I) encoding a LIN-8 polypeptide  
 CC (see ABB78648), LIN-56 polypeptide (see ABB78649), or LIN-61 polypeptide  
 CC (see ABB78650), where the polypeptides comprise at least 130, 116, 130  
 CC contiguous amino acids of the 386, 322 and 498 amino acid sequences given  
 CC in ABB78648 to ABB78650, and modulate cell proliferation. (I) has  
 CC cytostatic activity, and can be used in gene therapy. (I) is useful for  
 CC modulating proliferation of a cell, and for identifying the compound that  
 CC modulates cell proliferation. (I) can be used for diagnosing an animal  
 CC (preferably, human) for the presence of the cell proliferation disease,  
 CC or an increased chance of developing the disease, by measuring lin-8,  
 CC lin-56 or lin-61 nucleic acid expression in a sample obtained from the  
 CC animal, by measuring LIN-8, LIN-56 or LIN-61 polypeptide in the sample  
 CC using Ab or measuring amount of lin-8, lin-56, lin-61 mRNA in the sample.  
 CC (I) is also useful for identifying the compound that modulates cell  
 CC proliferation. (I) and its mutants are useful for understanding cell  
 CC proliferative diseases including cancer, as well as in diagnosing and  
 CC treating cell proliferative diseases. (I) is also useful for identifying  
 CC tumour suppressors in other species such as mammals and may be used to  
 CC identify therapeutic compounds. The present sequence represents a human  
 CC hsp (heat shock protein) protein which is given in comparison with the  
 CC C. elegans LIN-61 protein in the exemplification of the present  
 CC invention.  
 XX  
 XX Sequence 410 AA;  
 SQ  
 Query Match 51.7%; Score 1387; DB 23; Length 410;  
 Best Local Similarity 62.6%; Pred. No. 5.7e-135;  
 Matches 248; Conservative 56; Mismatches 84; Indels 8; Gaps 2;  
 QY 2 KGMKVELNSDAVLPSRVYVIASTVQAGYVLLRYEGFENDASHDFWNCNGTVDVHPIG 61  
 Db 10 ENVRVEVPNTDCLPTKFWIAGIYKLVKLTGAKTLPPDFSKVSESQYFPCMRVEV 69  
 QY 62 WCAINSKILVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESQYFPCMRVEV 121  
 Db 70 WCAASGKPLVPPRTTHAKFTDWKGYLMKRLVGSRTLPVDFFHKVSESQYFPCMRVEV 129

QY 122 VDKQVQRTMAVVDVTIGRLRLLYEDG-DSDDDFWCHMWSPLIHPVGSRRVGHGIM 180  
 DB 130 VDKHLCKRTRVAVSVESVIGRLRLLYVESDRTDDFWCHMWSPLIHPVGSRRVGHGIM 189  
 QY 181 SERSDMAHHTFRKIYCDVAVPLPKVRAVYVTEGGHFEFGMKLEADPLNLGNICVATV 240  
 DB 190 SDITKQDGHFT-----DPPHLFAKVKEVDQSGGEWFKGEMKLEADPLNLSTICVATI 242  
 QY 241 CKVLLDGLYLMICVDGSPSTDLGDFCYHASSHAIFPATFCQKNDIELTPPKGYEAOFTFW 300  
 DB 243 KRVLADGLMIGDSEADGSDWFCVHATSPSIFPVGCEINNIETLTPRGYTKLPFKW 302  
 QY 301 ENYLEKTSKAAPRELNMDCPAHGFVKVGMKLEADLMEPRLICVATVKEVVRLLSIHF 360  
 DB 303 FDLRETSSTAAPVKLENKQVFNHGFVGMKLEADLMEPRLICVATVTRIIHRLRIHF 362  
 QY 361 DGMDSYDQWDCESPDIYPVGMCELTGYQLQPPVA 396  
 DB 363 DGWEEYDQWDCESPDIYPVGMCELTGYQLQPPAS 398

RESULT 9

ID ABB70683 standard; Protein; 891 AA.

AC ABB70683;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 38841.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL14786.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

FT interactions -

XX Disclosure; SEQ ID NO 38841; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 891 AA;

Query Match 41.4%; Score 1111; DB 22; Length 891;

Best Local Similarity 43.4%; Pred. No. 9.4e-106;  
 Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;  
 QY 3 GKKVEVLNSD-----AVLPSRV---YMTASVQTAGVEVLRYEGFENDASHDFWCNLTGT 54  
 DB 218 GKKVEVENTDCDSIEVIQFGQTPTSFWVATILEIKGYKALMSYEGFDT-SHDFWVNLCN 276  
 QY 55 VDVHPIGHCAINSKILVPPRTIHAFTDKWGLYMKRLVGSRTLPVDHFHKVMSMKYPFR 114  
 DB 277 AEVHSVGCATRGKPLIPRTIEHKYKDWKDFLVGRSLSGARTLPSNFYKNKINDSLQSRFR 336  
 QY 115 QGMRLEVVDKQSVSTENAVVDVTIGRLRLLYEDGSDDDDFWCHMWSPLIHPVGSRRV 174  
 DB 337 LGNLEECVDKORISORLATVTIKVGRKFLRY--FDSDDGFWCHEDSPIIHPVGWATTV 394  
 QY 175 GHGI-----KMSERRSDMAHHTFRKIYCDVAVPLFK---KVRVAVYVTEG--GWPEE 220  
 DB 395 GHNLAAPODYLERMLAGREAMEVHEH-----DATIELFKMNTFFDEYVSDGKTNFVE 448  
 QY 221 GKKLEAIDPLNLGNICVATVCKVLLDGLYLMICVDG-GEPTDGLDWFCHASSHAIFPATF 279  
 DB 449 GKKLEAVDPLNLSSICPATVMAVLKFGYMMIRIDSYPDASGSDWFCYHEKSPCIFPAGF 508  
 QY 280 CQKNDIELTPPKGYEAOFTFWENYLEKTSKAAPSRLLFNMDCPNHGPKVGMKLEAVDLME 339  
 DB 509 CSVNNISVTPNGYDSRTFTWEGYLRDTGAVAAGQHLFHRIPDHGFVGMKLECADLMD 568  
 QY 340 PRLICVATVTKVRVRLLSIHFDGMDSEYDQWDCESPDIYPVGMCELTGYQLQPP----- 394  
 DB 569 PRLVCVATVARVVGRLKLVHFDGMDTDEYDQWDCESADIIYPVGMCELVNHHKLEGGPRVAH 628  
 QY 395 -VAEPATPLKAKEATKKKQFGKRRIPPTKTRPLRQSGKKPLLEDDEDDPGARKISSE 453  
 DB 629 QQAPKAPKPKIQKRRPKKGAAGK-----TPTDNTQSVKSRITALKTTTH----- 676  
 QY 454 PVPGEIIAVRVKKEHLDVA-----SPDKASSPELF-----VSVENIKQETDD 495  
 DB 677 -LPKLSIKLELKPHEHNAAFYENNQPEEGDEEDPDADGDGDGSGTSHISEQSTTQSSSD 734

RESULT 10

ID ABB66759 standard; Protein; 1243 AA.

AC ABB66759;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 27069.

KW Drosophila; developmental biology; cell signalling; insecticide;

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE ) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL10862.

PT New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell





QY 465 KEEHLDVAVP--DKASSPELPSV 486  
DB 506 YFNERCFSGPYLNKGRIAEFPQCV 529

RESULT 12

ABP69567  
ID ABP69567 standard; Protein; 894 AA.

AC ABP69567;  
XX

DT 20-JAN-2003 (first entry)  
XX

DE Human polypeptide SEQ ID NO 1614.  
XX

KW Human; genome mapping; gene therapy; food supplement; virus; fungus;  
cell-proliferative disorder; neurodegenerative disease; bacterial;  
Parkinson's disease; Alzheimer's disease; autoimmune disease;  
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;  
arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;  
antiParkinsonian; antidiabetic; immunosuppressive; dermatological;  
haemostatic; vulnary; fungicide; antibacterial; virucide; protozoacide;  
antiarthritic.

OS Homo sapiens.  
XX

PN WO200270539-A2.  
XX

PD 12-SEP-2002.  
XX

PF 05-MAR-2002; 2002WO-US05095.  
XX

PR 05-MAR-2001; 2001US-0799451.  
XX

FA (HYSE-) HYSEQ INC.  
XX

PI Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Ma Y, Yamazaki V, Chen R, Wang Z, Ghosh M;  
PI Wehrman T, Wang J, Wang D, Drmanac RT;

DR WPI; 2002-759812/82.  
DR N-PSDB; ABZ11784.

XX New polynucleotides comprising sequences assembled from expressed  
sequence tags (ESTs), useful for treating cell-proliferative,  
neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or  
platelet or coagulation disorders -

PS Claim 9; SEQ ID NO 1614; 1012pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (I) comprising a  
nucleotide sequence selected from any of 948 sequences  
(ABZ11119-ABZ12066) or their mature protein coding portion, active domain  
coding protein or complementary sequences. The polynucleotides are useful  
for identifying expressed genes or for physical mapping of human genome.  
The encoded polypeptides (ABP68902-ABP69849) are useful as molecular  
weight markers, as a food supplement, for generating antibodies, in  
medical imaging, screening and diagnostic assays and for treating  
cell-proliferative disorders (cancer), neurodegenerative diseases  
(Parkinson's or Alzheimer's disease), autoimmune diseases (multiple  
sclerosis, diabetes, lupus), genetic disorders, myeloid or lymphoid  
disorders, platelet or coagulation disorders, wound, burns, incision,  
ulcers, liver or lung fibrosis, infections (bacterial, viral, fungal,  
parasitic), arthritis, etc.

CC Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 894 AA;

Query Match 22.8%; Score 611; DB 23; Length 894;

Best Local Similarity 34.18; Pred. No. 8.8e-54;

Matches 156; Conservative 65; Mismatches 170; Indels 66; Gaps 15;

QY 3 GNMKEVLNSDAVLPSRVYMIASVIQTAGRYVLLRYEGFENDASHDFWNCNLTGTVVHPIGW 62  
DB 77 GNMKEVLANKN---NPDYVWATIIITTCGOLLRLRYCYGEDRRADFWCDVVIADLHPVGW 133  
QY 63 CAINSKILVPPRTIHAKEKTDWKGILMKELVGSRTLPVDFHIKMWESMKYKPPR----- 114  
DB 134 CTQNNKVLMPDPDAIKEKYTDWTEFLIRDLTGSRTAPANL-----LEGPLRGKGPIDLI 186  
QY 115 -QCMRELVVDKSVSRTRMAVVDTVIGRLRLLY---EDGSDDDDFWCHMSPPLIHPVGW 170  
DB 187 TVGSLIELODSQNPFPQYWIIVSVIENVGGLRLRYVGLDTSYDQ-WLFYLDYRLRPVGW 245  
QY 171 SRVGHGIMKERRSD-----MAHH--PFRKIYCDVY-----LFKVRVYTE 214  
DB 246 COE-----NKYRMDPPSEIYPLKMASEWKTLEKSLIDAAKFFELPMEVFKDHADLRS 298  
QY 215 GCMFEGMKLEAIDPLNLGNTCVATVCKVLGDYLMICVDG-GRSTDGLDFWCVHASSHA 273  
DB 299 --FFTVGKMLETVANMCEPFYISPASVTKVFNHFFQVITIDLRPEPSKSLMLC-HADSLG 355  
QY 274 IPPATFCQNDIELTPPKGYBAQTFNWNYLEKTSKAAAPSLRFLNMDCPNHGFKVGMKLE 333  
DB 356 ILPVQWCLKNGVSLTPPKGYSGQDFWADYHKHGAQEAAPPFCFRNTSFRGFTKMKLE 415  
QY 334 AVDLMEPRLCVATVYKRVVHRLLSIHFQGWSEYDQ-WVDCESPDYVGVGNCETGYQLQ 392  
DB 416 AVNPRNPGLCVASVSVKGRMLMLHLEGLQTPVPEVIVDVESMDIFFVGVCEANSY--- 472  
QY 393 PVVAAEPATPLKAKEAT---KKKKKQFGKKRRIPT 426  
DB 473 -----PLTAPHKTVSQKKRIAVVQPEKQLPPT 500

RESULT 13

ABG14398  
ID ABG14398 standard; Protein; 205 AA.

XX AC ABG14398;  
XX

DT 18-FEB-2002 (first entry)  
XX

DE Novel human diagnostic protein #14389.  
XX

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.  
XX

PN WO200175067-A2.  
XX

PD 11-OCT-2001.  
XX

XX 30-MAR-2001; 2001WO-US08631.  
XX

XX 31-MAR-2000; 2000US-0540217.  
XX

XX 23-AUG-2000; 2000US-0649167.  
XX

XX (HYSE-) HYSEQ INC.  
XX

XX Drmanac RT, Liu C, Tang YT;  
XX

XX WPI; 2001-639362/73.  
XX

XX N-PSDB; AAS78585.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX Claim 20; SEQ ID NO 44757; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 205 AA;

Query Match 21.9%; Score 587; DB 22; Length 205;  
Best Local Similarity 56.1%; Pred. No. 2.8e-52;  
Matches 11; Conservative 28; Mismatches 47; Indels 12; Gaps 3;  
QY 16 PSRYWIASVIQAGYVRLRYEGFENDASHDFWCLNLTGVDPVHPICWAINSKILVPPPT 75  
DB 14 PSRV---ALEQECGYNALLRYEGFENDSLDFWCLNLTGVDPVHPICWAINSKILVPPPT 69  
QY 76 IHAFTDKWYLMKRLVGSRTLPVDPHIKMVESMKYPRQGRLEVVVDSQVSRTRMAV 135  
DB 70 IQHNTNKAFLVRLTGAKTLPPDFPSQKVSSESNQYPRPCMKVEVVDKRLCRTRAV 129  
QY 136 DTVIGGLRLLYEDG-DSDDDFWCHWMSPLIHPVGSRRVGHGKIMSERRSDMAHPTFR 194  
DB 130 ESVGGLRLVYSESDRTDDFWCHWMSPLIHPVGHGKIMSERRSDMAHPTFR 185  
QY 195 KIYCDVAVPYLKKVAVY 212  
DB 186 ---FDTPPHLEFAKVSFY 200

RESULT 14  
ABU11746  
ID ABU11746 standard; Protein; 538 AA.  
XX  
AC ABU11746;

DT 13-FEB-2003 (first entry)  
XX  
DE Human MDDT polypeptide SEQ ID 693.

XX MDDT; human; disease detection and treatment molecule polypeptide;  
KW anti-inflammatory; immunosuppressive; osteoparatic; cytostatic; anti-HIV;  
KW haemostatic; nephrotropic; antianemic; antiparatic; hepatotropic;  
KW gene therapy; protein replacement therapy; cell proliferative disorder;  
KW cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;  
KW anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
KW Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopaenia;  
KW psoriasis; hepatitis.  
XX  
OS Homo sapiens.  
XX  
PN WO200279449-A2.  
XX  
PD 10-OCT-2002.  
XX  
XX 27-MAR-2002; 2002WO-US09944.  
PF  
XX 28-MAR-2001; 2001US-279619P.  
PR 29-MAR-2001; 2001US-280067P.

PR 29-MAR-2001; 2001US-280069P.  
PR 16-MAY-2001; 2001US-291280P.  
PR 17-MAY-2001; 2001US-291829P.  
PR 17-MAY-2001; 2001US-291849P.  
PR 19-JUN-2001; 2001US-299428P.  
PR 20-JUN-2001; 2001US-299776P.  
PR 20-JUN-2001; 2001US-300001P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;  
PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Ameshey SR;  
PI Daugherty SC, Dam TC, Liu TF, Nguyen DA, Klesfeld Y, Gerstin EH;  
PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;  
PI Flores V, Marwaha R, Lo A, Lan R, Urashka M;  
XX WPI; 2003-058431/05.  
XX N-PSDB; ABX34736.

XX New purified disease detection and treatment molecule proteins and  
PT polynucleotides, useful for diagnosing, treating or preventing cancers  
PT (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis  
PT or hepatitis  
XX  
XX Claim 27; SEQ ID NO 693; 339pp + Sequence Listing; English.

XX This invention describes a novel disease detection and treatment molecule  
CC polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
CC osteoparatic, cytostatic, anti-HIV, haemostatic, nephrotropic,  
CC antianemic, antiparatic and hepatotropic activity. The polynucleotides  
CC and the polypeptides of the invention can be used for gene therapy,  
CC protein replacement therapy and are useful for treating a variety of  
CC diseases or conditions. These polypeptides or polynucleotides are  
CC particularly useful for diagnosing, treating or preventing cell  
CC proliferative disorders (e.g. cancers including adenocarcinoma,  
CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
CC syndromes, inflammation, osteoporosis, thrombocytopaenia, psoriasis or  
CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded  
CC by ABU11450-ABU11845, described in the disclosure of the invention.  
CC NOTE: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 538 AA;

Query Match Similarity 19.0%; Score 510; DB 24; Length 538;  
Best Local Similarity 27.4%; Pred. No. 1.3e-43;  
Matches 132; Conservative 50; Mismatches 144; Indels 156; Gaps 12;

QY 3 GMKVEVLNSDAVLPSPRYWIASVIQAGYVRLRYEGFENDASHDFWCLNLTGVDPVHPIGW 62  
DB 27 GMKLEGI--DPQHPS-MYFILTVAEVCYGLRLHFDGYS--CHDFWVWNSPDIHPAGW 81  
QY 63 CAINSKILVPPRTTHAKFTDWKGMKRLVGSRTLPVDPHIKMVESMKYP---FQGMRL 119  
DB 82 FEKTGHKLQPKGYKEEFSSQYLR---STRAQAAPKELFVSQSHPPLGFQVQMKL 137  
QY 120 EVDKQSQRTRMAVDPVIGGRLLVYEDGSDDDDFWCHWMSPLIHPVGSRRVGHGK 179  
DB 138 EAVDRMNPFLVCASVTDVDSRLVHFDNWDYDTCDFPSSYIHPVGW----- 188  
QY 180 MSERRSDMAHPTFRKIYCDVAVPYLKKVAVYTEGGWFEGMKLEAIDPLNLGNICVAT 239  
DB 189 ----- 188  
QY 240 VKVLLDGYLMICVDGGPSTGLDWFVCHASSHAIFPATFCKNDIELTPPKGY-EACTF 298  
DB 189 -----CQCKEKLTPPDYDFDNF 208  
QY 299 NWENYLEKTSKAAAPSRFLFNMDCPNHGFKVGMKLEAVDLMEPRLICVATKRVVHLLSI 358  
DB 209 CWKYLEETGASAVPTWAFKVR--PPHSLVNMKLEAVDRRNPAIRVASVEDVEDHRIKI 267

QY 359 HFDGWDSEYDQWDCESPDYIPVGVKCELTGYQLQPPVA-AEPAT-----PLKAKEATK 410  
 D 268 HFDGWSHGVDWIDADHPDIHPACWCKTGHPLOPLGPREPSSASPGGCPPLSYRSLPH 327  
 QY 411 KKKQFGKKRKRIP-----PTKTRPLRQSKKPLEDDPQGA 447  
 D 328 TRTSKYSFHHKCKPTPGCDGSGHVTGKFTAHHCLSGCPFLAER--NQSLKAEKLSDEASA 385  
 QY 448 RK 449  
 D 386 RK 387

RESULT 15  
 ABU11833  
 ID ABU11833 standard; Protein; 538 AA.  
 AC ABU11833;  
 XX

13-FEB-2003 (first entry)

Human MDDT polypeptide SEQ ID 780.

MDDT: human; disease detection and treatment molecule polypeptide;  
 anti-inflammatory; immunosuppressive; osteoporosis; cytostatic; anti-HIV;  
 haemostatic; nephrotropic; antianaemic; antiproliferative; hepatotropic;  
 gene therapy; protein replacement therapy; cell proliferative disorder;  
 cancer; adenocarcinoma; leukaemia; lymphoma; melanoma; myeloma; sarcoma;  
 anaemia; Crohn's disease; acquired immunodeficiency syndrome; AIDS;  
 Goodpasture's syndrome; inflammation; osteoporosis; thrombocytopenia;  
 psoriasis; hepatitis.

XX Homo sapiens.

XX WO200279449-A2.

XX 10-OCT-2002.

XX 27-MAR-2002; 2002WO-US09944.

XX 28-MAR-2001; 2001US-279619P.

XX 29-MAR-2001; 2001US-280067P.

XX 29-MAR-2001; 2001US-280069P.

XX 16-MAY-2001; 2001US-291280P.

XX 17-MAY-2001; 2001US-291829P.

XX 17-MAY-2001; 2001US-291849P.

XX 19-JUN-2001; 2001US-299428P.

XX 20-JUN-2001; 2001US-299776P.

XX 20-JUN-2001; 2001US-300001P.

XX (INCY-) INCYTE GENOMICS INC.

XX Dafo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;

XX Dufour GE, Hillman JL, Yu JY, Tauson O, Yap PE, Anshey SR;

XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;

XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;

XX Flores V, Marwaha R, Lo A, Lan RY, Urashka MS;

XX WPI; 2003-058431/05.

XX N-FSDB; ABX34823.

XX

XX New purified disease detection and treatment molecule proteins and

XX polynucleotides, useful for diagnosing, treating or preventing cancers

CC and the polypeptides of the invention can be used for gene therapy,  
 CC protein replacement therapy and are useful for treating a variety of  
 CC diseases or conditions. These polypeptides or polynucleotides are  
 CC particularly useful for diagnosing, treating or preventing cell  
 CC proliferative disorders (e.g. cancers including adenocarcinoma,  
 CC leukaemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's  
 CC disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's  
 CC syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or  
 CC hepatitis. ABU11450-ABU11845 represent the MDDT polynucleotides encoded  
 CC by ABU11450-ABU11845, described in the disclosure of the invention.  
 CC NOTE: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 538 AA;

Query Match 19.0%; Score 510; DB 24; Length 538;

Best Local Similarity 27.4%; Pred. No. 1.3e-43;

Matches 132; Conservative 50; Mismatches 144; Indels 156; Gaps 12;

QY 3 GMKVEVLNSDAVLPSSRVYVIASVIQTAGYRVLLYEGENDASHDFWNLGTVDVHPGW 62

DB 27 GMKLEGI--DPCHPS--MYFILTVAEVCYRLRLHFDGYSE--CHDFWVNANSPDIHPAGW 81

QY 63 CAINSKILVPRTIHAKEFTDKGVMKELVGSRTLVPDFFHKMYESMKYP---FQGNRL 119

DB 82 FEKTHKLPQPKGYKEEFSSQYLK---STRAQAAPKHLFVSQSHSPPLGFQVGMKL 137

QY 120 EYVDKSVSRTRMAVDTVIGRLRLLYEDGSDDDDFWCHMWSPLIHPVGSRRVHGKIK 179

DB 138 EAVDRMNSLVCVASVTDVDSRLVHFDNWDYDWCDFSSPYIHPVGW----- 188

QY 180 MSERSDMAHPTFRKIYCDAPVPLFKKRVAYTTEGGWFEGMKLEAIDPLNLGNCVAT 239

DB 189 ----- 188

QY 240 VCKVLLDGYLMICVDGGSPSTGLDWFVCHASSHAIFPATFCQKNDIELTPPKGY-EACTF 298

DB 189 -----CQKQKPLTPPDYFDPDNF 208

QY 299 NWNYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATKRVVHRLLSI 358

DB 209 CWKYLEETGASAVPTWAFKVR--PPHSFLVNNKLEAVDRRNPALIRVASVEDVEDHRIKI 267

QY 359 HFDGWDSEYDQWDCESPDYIPVGVKCELTGYQLQPPVA-AEPAT-----PLKAKEATK 410

DB 268 HFDGWSHGVDWIDADHPDIHPACWCKTGHPLOPLGPREPSSASPGGCPPLSYRSLPH 327

QY 411 KKKQFGKKRKRIP-----PTKTRPLRQSKKPLEDDPQGA 447

DB 328 TRTSKYSFHHKCKPTPGCDGSGHVTGKFTAHHCLSGCPFLAER--NQSLKAEKLSDEASA 385

QY 448 RK 449

DB 386 RK 387

Search completed: February 4, 2004, 09:06:54

Job time : 49 secs

Claim 27; SEQ ID NO 780; 339pp + Sequence Listing; English.

This invention describes a novel disease detection and treatment molecule  
 polypeptide (MDDT) which has anti-inflammatory, immunosuppressive,  
 osteoporotic, cytostatic, anti-HIV, haemostatic, nephrotropic,  
 antianaemic, antiproliferative and hepatotropic activity. The polynucleotides



Db 67 GNCENKGMQLPPLGFRNLNASSWPMFLKTLNGAEMAPIRIFHKBPSPSHNFFKMGKML 126  
Qy 333 EAVDLMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDCSPDIYPVGMWCELGYQLQ 392  
Db 127 EAVDRKNPHFICPATIGEVRGSEVLVTFDGRGAFDYWCFRDSRDIFFVGMWCSLGDNIQ 186  
Qy 393 PP--VAEPATPLKAEATKKK-----KQFGKKRKRIP----- 424  
Db 187 PPGTKVVPKPNYPASDVNTEKPSIHSSTKTVLEHQPQGRKPKGKGRTPKTLISHPI 246  
Qy 425 --PTKT-RPL-----RQSK-KPLEDDPOGARKISSEVPV 456  
Db 247 SAPSKTAEPKPKGPKGSKRKRTLLNPPPPASPTTSTPEP 290

RESULT 2  
US-08-852-153-2  
; Sequence 2, Application US/08852153  
; Patent No. 5914266  
; GENERAL INFORMATION:  
; APPLICANT: Randazzo, Filippo  
; TITLE OF INVENTION: Mammalian Sex Comb on Midleg Acts as a Tumor Suppressor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,153  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guth, Joseph H.  
; REGISTRATION NUMBER: 31,261  
; REFERENCE/DOCKET NUMBER: 1224.006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-3888  
; TELEFAX: (510) 555-3542  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 620 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-852-153-2

Query Match 13.3%; Score 357.5; DB 2; Length 620;  
Best Local Similarity 35.9%; Pred. No. 4.5e-31;  
Matches 102; Conservative 31; Mismatches 102; Indels 49; Gaps 11;

Qy 218 FEEGMKLEADPLNLGNICVATCKVLLDGYLMICVDGSPSTGLDWFCHASSHAIFPA 277  
Db 40 FKISMKLEAQDPRNTTSTCIATVVG-LTGARLRLDG---SDNKDFWRLVDSAEIQPI 95  
Qy 278 TFCQNDIELTPPKGYEAQTFNWNYLEKTKSKA--APSRLENMD--CPNHG-FKVGKML 332  
Db 96 GNCENKGMQLPPLGFRNLNASSWPMFLKTLNGAEMAPIRIFHKBPSPSHNFFKMGKML 155  
Qy 333 EAVDLMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDCSPDIYPVGMWCELGYQLQ 392  
Db 156 EAVDRKNPHFICPATIGEVRGSEVLVTFDGRGAFDYWCFRDSRDIFFVGMWCSLGDNIQ 215  
Qy 393 PP--VAEPATPLKAEATKKK-----KQFGKKRKRIP----- 424

Db 216 PPGTKVVPKPNYPASDVNTEKPSIHSSTKTVLEHQPQGRKPKGKGRTPKTLISHPI 275  
Qy 425 --PTKT-RPL-----RQSK-KPLEDDPOGARKISSEVPV 456  
Db 276 SAPSKTAEPKPKGPKGSKRKRTLLNPPPPASPTTSTPEP 319

RESULT 3  
US-08-852-153-4  
; Sequence 4, Application US/08852153  
; Patent No. 5914266  
; GENERAL INFORMATION:  
; APPLICANT: Randazzo, Filippo  
; TITLE OF INVENTION: Mammalian Sex Comb on Midleg Acts as a Tumor Suppressor  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Chiron Corporation  
; STREET: 4560 Horton Street  
; CITY: Emeryville  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94608  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/852,153  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Guth, Joseph H.  
; REGISTRATION NUMBER: 31,261  
; REFERENCE/DOCKET NUMBER: 1224.006  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (510) 923-3888  
; TELEFAX: (510) 555-3542  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-852-153-4

Query Match 13.2%; Score 353.5; DB 2; Length 577;  
Best Local Similarity 36.1%; Pred. No. 1.1e-30;  
Matches 101; Conservative 30; Mismatches 100; Indels 49; Gaps 11;

Qy 222 MKLEAIDPLNLGNICVATCKVLLDGYLMICVDGSPSTGLDWFCHASSHAIFPAFCQ 281  
Db 1 MKLEAQDPRNTTSTCIATVVG-LTGARLRLDG---SDNKDFWRLVDSAEIQPENCE 56  
Qy 282 KNDIELTPPKGYEAQTFNWNYLEKTKSKA--APSRLENMD--CPNHG-FKVGKMLEAVD 336  
Db 57 KNGGMQLPPLGFRNLNASSWPMFLKTLNGAEMAPIRIFHKBPSPSHNFFKMGKMLEAVD 116  
Qy 337 LMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDCSPDIYPVGMWCELGYQLQPP-- 394  
Db 117 RKNPHFICPATIGEVRGSEVLVTFDGRGAFDYWCFRDSRDIFFVGMWCSLGDNIQPPGT 176  
Qy 395 VAEPATPLKAEATKKK-----KQFGKKRKRIP-----PT 426  
Db 177 KVVIPKPNYPASDVNTEKPSIHSSTKTVLEHQPQGRKPKGKGRTPKTLISHPIAPS 236  
Qy 427 KT-RPL-----RQSK-KPLEDDPOGARKISSEVPV 456  
Db 237 KTASPLPKPKGPKGSKRKRTLLNPPPPASPTTSTPEP 276

RESULT 4

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:08:22 ; Search time 39 Seconds

(without alignments)  
2657.544 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKEVLNSDAVLPSRV.....KASSPLPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 801455 seqs, 209382283 residues

Total number of hits satisfying chosen parameters: 801455

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	2679	99.9	705	12	US-10-094-749-2556
2	2185	81.4	526	12	US-10-108-260A-4101
3	1688	62.9	329	12	US-10-264-237-1824
4	1410	52.6	331	9	US-09-789-919-56
5	1387	51.7	410	10	US-09-872-523-68
6	495.5	18.5	623	12	US-10-104-047-3512
7	494.5	18.4	498	10	US-09-872-523-5
8	493.5	18.4	498	10	US-09-872-523-71
9	486.5	18.1	498	10	US-09-872-523-72
10	440	16.4	78	12	US-10-029-386-27484
11	417	15.5	77	9	US-09-864-761-38826
12	417	15.5	77	12	US-10-029-386-30438
13	398	14.8	427	12	US-10-104-047-2262
14	389.5	14.5	406	12	US-10-104-047-3204
15	373.5	13.9	217	12	US-10-264-049-2539

16	340	12.7	827	9	US-09-801-574-12
17	330.5	12.3	100	10	US-09-872-523-7
18	302.5	11.3	411	10	US-09-872-523-70
19	281.5	10.5	125	15	US-10-204-887-93
20	272	10.1	512	10	US-09-872-523-69
21	271.5	10.1	96	10	US-09-872-523-10
22	268.5	10.0	100	10	US-09-872-523-9
23	259.5	9.7	186	12	US-10-104-047-2159
24	242.5	9.0	311	10	US-09-764-868-1009
25	238.5	8.9	100	10	US-09-872-523-12
26	236.5	8.8	100	10	US-09-872-523-8
27	233.5	8.3	100	10	US-09-872-523-15
28	211	7.9	99	10	US-09-872-523-11
29	199	7.4	38	9	US-09-864-761-38862
30	186.5	7.0	96	10	US-09-872-523-13
31	178	6.6	99	10	US-09-872-523-14
32	177	6.6	67	9	US-09-864-761-36870
33	162	6.0	78	9	US-09-864-761-48140
34	158	5.9	74	12	US-10-029-386-33605
35	141	5.3	25	9	US-09-864-761-42641
36	122	4.5	25	9	US-09-864-761-38882
37	111	4.1	790	12	US-10-369-493-3803
38	102	3.8	512	12	US-10-320-797-3205
39	94	3.5	634	12	US-10-369-493-13304
40	93.5	3.5	371	9	US-09-895-943-5
41	93.5	3.5	371	10	US-09-895-593-5
42	93.5	3.5	371	11	US-09-376-430-2
43	93.5	3.5	371	12	US-10-354-951-3
44	93.5	3.5	371	14	US-10-008-566-4
45	93.5	3.5	371	14	US-10-078-059-2

# ALIGNMENTS

RESULT 1  
US-10-094-749-2556  
; Sequence 2556, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:

- ; APPLICANT: ISOGAI, TAKAO
- ; APPLICANT: SUGIYAMA, TOMOYASU
- ; APPLICANT: OTSUKI, TETSUJI
- ; APPLICANT: WAKAMATSU, AI
- ; APPLICANT: SATO, HIROYUKI
- ; APPLICANT: ISHII, SHIZUKO
- ; APPLICANT: YAMAMOTO, JUN-ICHI
- ; APPLICANT: ISONO, YUUKO
- ; APPLICANT: HIO, YURI
- ; APPLICANT: OTSUKA, KAORU
- ; APPLICANT: NAGAI, KEIICHI
- ; APPLICANT: IRIE, RYOTARO
- ; APPLICANT: TAMECHIKA, ICHIRO
- ; APPLICANT: SEKI, NACHIKO
- ; APPLICANT: YOSHIKAWA, TSUTOMU
- ; APPLICANT: OTSUKA, MOTOTYUKI
- ; APPLICANT: NAGAHARI, KENJI
- ; APPLICANT: MASUHO, YASUHIKO
- ; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
- ; FILE REFERENCE: 084335/0160
- ; CURRENT APPLICATION NUMBER: US/10/094,749
- ; CURRENT FILING DATE: 2002-03-12
- ; PRIOR APPLICATION NUMBER: 60/350,435
- ; PRIOR FILING DATE: 2002-01-24
- ; PRIOR APPLICATION NUMBER: JP 2001-328381
- ; PRIOR FILING DATE: 2001-09-14
- ; NUMBER OF SEQ ID NOS: 3381
- ; SOFTWARE: PatentIn Ver. 2.1
- ; SEQ ID NO 2556
- ; LENGTH: 705
- ; TYPE: PRT
- ; ORGANISM: Homo sapiens

US-10-094-749-2556

Query Match 99.9%; Score 2679; DB 12; Length 705;  
 Best Local Similarity 99.8%; Pred. No. 4.5e-259;  
 Matches 494; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKGKVEVLSNDVLPSSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60  
 DB 211 MKGKVEVLSNDVLPSSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 270  
 QY 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKRLVGSRTLPVDFHFKMVESMKYPFQGMRL 120  
 DB 271 GWCAINSKILVPPRTTHAKFTDMKGYLMKRLVGSRTLPVDFHFKMVESMKYPFQGMRL 330  
 QY 121 VDKSQSVSRTRMAVVDVTFVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 180  
 DB 331 VDKSQSVSRTRMAVVDVTFVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 390  
 QY 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEAIDPLNLGNICVATV 240  
 DB 391 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEAIDPLNLGNICVATV 450  
 QY 241 CKVLLDGYLMICVDGGSPDGLDFWCFYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300  
 DB 451 CKVLLDGYLMICVDGGSPDGLDFWCFYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 510  
 QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIHF 360  
 DB 511 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIHF 570  
 QY 361 DGWDSYDQWDCSPDIYPVGMCELTGYQLQPPVAAEPATPLFAEATKKKKQFGKKR 420  
 DB 571 DGWDSYDQWDCSPDIYPVGMCELTGYQLQPPVAAEPATPLFAEATKKKKQFGKKR 630  
 QY 421 KRIPPTKTRPLROGSKPILLEDPOGARKISSEFPVGEIIAVRVKSEHLDVASPKKASP 480  
 DB 631 KRIPPTKTRPLROGSKPILLEDPOGARKISSEFPVGEIIAVRVKSEHLDVASPKKASP 690  
 QY 481 ELPVSVENIKQETDD 495  
 DB 691 ELPVSVENIKQETDD 705

RESULT 2  
 US-10-108-260A-4101  
 ; Sequence 4101, Application US/10108260A  
 ; Publication No. US20040005560A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna  
 ; FILE REFERENCE: HI-A0106  
 ; CURRENT APPLICATION NUMBER: US/10/108,260A  
 ; PRIOR FILING DATE: 2002-03-27  
 ; NUMBER OF SEQ ID NOS: 5458  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 4101  
 ; LENGTH: 526  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-108-260A-4101

Query Match 81.4%; Score 2185; DB 12; Length 526;  
 Best Local Similarity 100.0%; Pred. No. 8.9e-210;  
 Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKGKVEVLSNDVLPSSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60  
 DB 120 MKGKVEVLSNDVLPSSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 179  
 QY 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKRLVGSRTLPVDFHFKMVESMKYPFQGMRL 120  
 DB 180 GWCAINSKILVPPRTTHAKFTDMKGYLMKRLVGSRTLPVDFHFKMVESMKYPFQGMRL 239  
 QY 121 VDKSQSVSRTRMAVVDVTFVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 180

DB 240 VDKSQSVSRTRMAVVDVTFVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 299  
 QY 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEAIDPLNLGNICVATV 240  
 DB 300 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEAIDPLNLGNICVATV 359  
 QY 241 CKVLLDGYLMICVDGGSPDGLDFWCFYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300  
 DB 360 CKVLLDGYLMICVDGGSPDGLDFWCFYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 419  
 QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIHF 360  
 DB 420 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVKRVVHRLLSIHF 479  
 QY 361 DGWDSYDQWDCSPDIYPVGMCELTGYQLQPPVAA 397  
 DB 480 DGWDSYDQWDCSPDIYPVGMCELTGYQLQPPVAA 516

RESULT 3  
 US-10-264-237-1824  
 ; Sequence 1824, Application US/10264237  
 ; Publication No. US20040009491A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birse et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PAL3P1  
 ; CURRENT APPLICATION NUMBER: US/10/264,237  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: PCT/US01/16450  
 ; PRIOR FILING DATE: 2001-05-18  
 ; PRIOR APPLICATION NUMBER: US 60/205,515  
 ; PRIOR FILING DATE: 2000-05-19  
 ; NUMBER OF SEQ ID NOS: 2876  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 1824  
 ; LENGTH: 329  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE: MISC FEATURE  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (185)  
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
 ; NAME/KEY: MISC FEATURE  
 ; LOCATION: (223)  
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
 ; US-10-264-237-1824

Query Match 62.9%; Score 1688; DB 12; Length 329;  
 Best Local Similarity 99.0%; Pred. No. 2.8e-160;  
 Matches 308; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MKGKVEVLSNDVLPSSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60  
 DB 1 MKGKVEVLSNDVLPSSRVVWIASVIQTAGYRVLLRYEGFENDASHDFWNLGTVDVHPI 60  
 QY 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKRLVGSRTLPVDFHFKMVESMKYPFQGMRL 120  
 DB 61 GWCAINSKILVPPRTTHAKFTDMKGYLMKRLVGSRTLPVDFHFKMVESMKYPFQGMRL 120  
 QY 121 VDKSQSVSRTRMAVVDVTFVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 180  
 DB 121 VDKSQSVSRTRMAVVDVTFVIGRLRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGIM 180  
 QY 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEAIDPLNLGNICVATV 240  
 DB 181 SERRSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEGMKLEAIDPLNLGNICVATV 240  
 QY 241 CKVLLDGYLMICVDGGSPDGLDFWCFYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300  
 DB 241 CKVLLDGYLMICVDGGSPDGLDFWCFYHASSHAIFPATFCQKNDIELTTPKGYEAQTFNW 300



```

QY 301 ENYLEKTSKA 311
Db 301 ENYLEKTSKA 311

RESULT 4
US-09-789-919-56
; Sequence 56, Application US/09789919
; Patent No. US20020064855A1
; GENERAL INFORMATION:
; APPLICANT: Lemischka, Ibor
; APPLICANT: Moore, Kateri
; TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM
; TITLE OF INVENTION: CELLS AND USES THEREOF
; FILE REFERENCE: 2275-1-005
; CURRENT APPLICATION NUMBER: US/09/789,919
; CURRENT FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 56
; LENGTH: 631
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-789-919-56

Query Match 52.6%; Score 1410; DB 9; Length 631;
Best Local Similarity 56.2%; Pred. No. 5.1e-132;
Matches 263; Conservative 67; Mismatches 122; Indels 16; Gaps 4;

QY 2 KGMKEVLENSDAVLPSRVYTIASVIQTAGYRVLLRYEGFENDASHDFWNCNLGTVDVHPG 61
Db 177 ENVAIEVNTDCSLPTKVFVWAGIAGVLLRYEGFENDASHDFWNCNLGTVDVHPG 236
QY 62 WCAINSKILVPPRTTHAKFTDWKYLKRLVGSRTLPVDPHIKVSMKYPFQGMRLV 121
Db 237 WCAASGKPLVPPRTVQHKYTNWKAFLVKRLTGAKTLPPDFSQKVSQSMQYFPCMRVEV 296
QY 122 VDKSOVSRTRMAVVDVTIGRLLLYEDG-DSDDDFCHMWSPLIHPVGSRRVGHGK 180
Db 297 VDKHLCRTVAVVESVIGRLLRYEESDGDGDFCHMWSPLIHPVGSRRVGHGK 356
QY 181 SERRSDMAHPTFRKIYCDAPVYLFKKVRAVYTEGGFEEGKLEAIDPLNLGNCVATV 240
Db 357 SDITPKQDGH-----FDTPPHLFAKVEVDQSGEWFKEGKLEAIDPLNLSTICVATI 409
QY 241 CKVLLDGLYLMICVDGPGSTGDLWFCYHASSHAIFPATPCOKNDIELTPPKGYEAQTFNW 300
Db 410 RKVLADGFLMIGIDGSEADGSDWFCYHATSPSIFPVGFCFCEINMIELTPPRGYTKLPFKW 469
QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVVRVHLLSIHF 360
Db 470 FDIYLETGSIAPVKLFNKDVPNHGFRVGMKLEAVDLMEPRLICVATVTRIHLRLIHF 529
QY 361 DGMSEYDQWDCSPDIYPVGMCELTGYQLQPPVAAEAPATPKAKKATKK-KKQFGKK 419
Db 530 DGMSEYDQWDCSPDLNPNVGMCELTGYQLQPPASQSSRESQSSAKSKAKSKQYKG 589
QY 420 RKPIPTKTRPLQSGKKPLEDDPGARKISSEPVGTEIIAVRKEE 467
Db 590 HKKWTTSQ-----XBEILLDGEDYSFLHGASDQSGNSGATVYIKOE 630

RESULT 5
US-09-782-523-68
; Sequence 68, Application US/09872523
; Patent No. US20020137906A1
; GENERAL INFORMATION:
; APPLICANT: Horvitz, H. Robert
; APPLICANT: Davison, Ewa M.
; APPLICANT: Lu, Xiaowei
; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans
; FILE REFERENCE: 01997/536002

```

```

; CURRENT APPLICATION NUMBER: US/09/872,523
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: US 60/208,802
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 68
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-872-523-68

Query Match 51.7%; Score 1387; DB 10; Length 410;
Best Local Similarity 62.6%; Pred. No. 5.4e-130;
Matches 248; Conservative 56; Mismatches 84; Indels 8; Gaps 2;

QY 2 KGMKEVLENSDAVLPSRVYTIASVIQTAGYRVLLRYEGFENDASHDFWNCNLGTVDVHPG 61
Db 10 ENVRVEVPNTDCSLPTKVFVWAGIAGVLLRYEGFENDASHDFWNCNLGTVDVHPG 69
QY 62 WCAINSKILVPPRTTHAKFTDWKYLKRLVGSRTLPVDPHIKVSMKYPFQGMRLV 121
Db 70 WCAASGKPLVPPRTIQQHKYTNWKAFLVKRLTGAKTLPPDFSQKVSQSMQYFPCMRVEV 129
QY 122 VDKSOVSRTRMAVVDVTIGRLLLYEDG-DSDDDFCHMWSPLIHPVGSRRVGHGK 180
Db 130 VDKHLCRTVAVVESVIGRLLRYEESDGDGDFCHMWSPLIHPVGSRRVGHGK 189
QY 181 SERRSDMAHPTFRKIYCDAPVYLFKKVRAVYTEGGFEEGKLEAIDPLNLGNCVATV 240
Db 190 SDITPKQDGH-----DPPHLFAKVEVDQSGEWFKEGKLEAIDPLNLSTICVATI 242
QY 241 CKVLLDGLYLMICVDGPGSTGDLWFCYHASSHAIFPATPCOKNDIELTPPKGYEAQTFNW 300
Db 243 RKVLADGFLMIGIDGSEADGSDWFCYHATSPSIFPVGFCFCEINMIELTPPRGYTKLPFKW 302
QY 301 ENYLEKTSKAAPSRFLNMDCPNHGFKVGMKLEAVDLMEPRLICVATVVRVHLLSIHF 360
Db 303 FDIYLETGSIAPVKLFNKDVPNHGFRVGMKLEAVDLMEPRLICVATVTRIHLRLIHF 362
QY 361 DGMSEYDQWDCSPDIYPVGMCELTGYQLQPPVA 396
Db 363 DGMSEYDQWDCSPDLNPNVGMCELTGYQLQPPAS 398

RESULT 6
US-10-104-047-3512
; Sequence 3512, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 3512
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3512

Query Match 18.5%; Score 495.5; DB 12; Length 623;
Best Local Similarity 36.6%; Pred. No. 2.1e-40;
Matches 121; Conservative 53; Mismatches 118; Indels 39; Gaps 12;

QY 83 WKGYLMKRLVGSRTLPVDPHIKVSMKYP-----FQGMRLVWVDSQVSRTEMAVVD 137
Db 54 WEWYKLEQ--KAAAPVELFSK---DQSFPEHENGFOIGRLLEGIDPRHFSVFCVLSVAE 108

```

Qy 138 VIGRLRLLYEDGSDDDDFWCHWMSPLIHPVGRSRRVGHGIMKERRSDMAHPT-PRK- 195  
Db 109 VCGYRLRLHFDGLSCYDFWNTAGSDIHPVGCERTKHEL-----HIPKGYRKO 158  
Qy 196 --IYCD-----AVPYLFKVRVAVYTEGGWFEQMKLEAIDPLNLGNICVATVCKVL 244  
Db 159 KFWMDYLKACKLQNAFKLFRNRPNGPMSEKFEQVGMKLEAVDRKNFSLVCVATIAIV 218  
Qy 245 LDGLYLMICVDGSGPSTDLGWFCHASHAIFPATFCQKNDIELTPPKGY-EAQTENWENY 303  
Db 219 ED-RLILVHFDNW--DSDYDWC--DVNSPYQPVGWCEQNGRTLAPQGYPENESWTEY 274  
Qy 304 LEKTKKAAPSRLENMDCNPHGKGMKLEAIDPLNLGNICVATVCKVVRHLLSHFGCW 363  
Db 275 LEATQTNVAPKVKRRLP-HGFLPNMKLEVDKRNPRILRVATIVDQDQKRVKHFPGW 333  
Qy 364 DSEYDWDCESDPIYPVGCWELTGYOLOPP 394  
Db 334 DHKYDYTWVADSPIHPIGWCDVTGHPLEVP 364

## RESULT 7

US-09-872-523-5  
; Sequence 5, Application US/09872523  
; Patent No. US20020137906A1  
; GENERAL INFORMATION:  
; APPLICANT: Horvitz, H. Robert  
; APPLICANT: Davison, Ewa M.  
; APPLICANT: Lu, Xiaowei  
; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans  
; FILE REFERENCE: 01997/536002  
; CURRENT APPLICATION NUMBER: US/09/872,523  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: US 60/208,802  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 498  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-872-523-5

Query Match 18.4%; Score 494.5; DB 10; Length 498;  
Best Local Similarity 29.7%; Pred. No. 1.9e-40;  
Matches 132; Conservative 81; Mismatches 149; Indels 83; Gaps 18;  
Qy 10 NSDAVLPSPRYVYIASVIQTAGYRVLRYEGFENDASHDFWCMNLGTVDVHPICWAINSK- 68  
Db 72 NCDSDI---QVRWFARIEKVCGRVLAQFIG---ADTKFWLNLSDDMFGLANAAMSDPN 124  
Qy 69 ----ILVPPRTIHAKE-TWKGYLMK----RLVGSRTLVDVDFHIKMWESMKYPPFGMRL 119  
Db 125 MDKIVVAPPLAINEEYQNDMVNVCIDGIEVGTSLSPKFDGKALLSKHRFKVGQRL 184  
Qy 120 EVDKSCQVSRTRNAVVDVTVIGRLR--LLYED-----GSDDDD-----FWCHMWSP 163  
Db 185 ELLNYSNSTEIRVARIEQICGRMNVSIITKDPFESLPDADDDRVFSSGSQYWIDEGSF 244  
Qy 164 LIHPVGM-----SRVGHGIMKERRSDMAHPTFRKIYCDV-PYLF 205  
Db 245 FIFPVGFAAVNGYQLNKAKKEYIEHTNKIAQAKNGENPRYDSDVTFDQAKDPIDPMIW 304  
Qy 206 KKVAVYTEGGWFEQMKLEAIDPL--NLGNICVATV--CKVLLDGYLMICVDGPGSTD 260  
Db 305 RKVKV-----GQKPELIDPLAQQFNHLVASILKFKCT--EGYLIVGMDGPDAL 352  
Qy 261 GLDMFCYCHASHAIFPATFCQKNDIELTPPKGYEACTFNWENYLEKTKSKAAPSRLFNMD 320  
Db 353 --DSFPIHINNTFMFVGYABKYNLELVPPDFKGTFRWDEYLEKESAEITLPLDLFK-P 408  
Qy 321 CPNH-----GFKVGMKLEAVDLMPEPLICVATVCKVVRHLLSHFGDWDSEYDQ 369  
Db 364 DSEYDWDCESDPIYPVGCWELTGYOLOPP 394  
Db 334 DHKYDYTWVADSPIHPIGWCDVTGHPLEVP 364

## RESULT 9

US-09-872-523-72  
; Sequence 72, Application US/09872523  
; Patent No. US20020137906A1  
; GENERAL INFORMATION:  
; APPLICANT: Horvitz, H. Robert  
; APPLICANT: Davison, Ewa M.

Db 409 MPQERLDKFKVLLISKVGLRLEAADMCENQFICPATVKSVHGRLLNVNFDGWDSEFDE 468  
Qy 370 WDCESPDYIPVGCWELTGYOLOPP 394  
Db 469 LYVDVSDHILPIGWCEAHSYVLOPP 493  
RESULT 8  
US-09-872-523-71  
; Sequence 71, Application US/09872523  
; Patent No. US20020137906A1  
; GENERAL INFORMATION:  
; APPLICANT: Horvitz, H. Robert  
; APPLICANT: Davison, Ewa M.  
; APPLICANT: Lu, Xiaowei  
; TITLE OF INVENTION: A Tumor Suppressor Pathway in C. Elegans  
; FILE REFERENCE: 01997/536002  
; CURRENT APPLICATION NUMBER: US/09/872,523  
; CURRENT FILING DATE: 2001-06-01  
; PRIOR APPLICATION NUMBER: US 60/208,802  
; PRIOR FILING DATE: 2000-06-02  
; NUMBER OF SEQ ID NOS: 78  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 71  
; LENGTH: 498  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-09-872-523-71

Query Match 18.4%; Score 493.5; DB 10; Length 498;  
Best Local Similarity 29.7%; Pred. No. 2.4e-40;  
Matches 132; Conservative 81; Mismatches 149; Indels 83; Gaps 18;  
Qy 10 NSDAVLPSPRYVYIASVIQTAGYRVLRYEGFENDASHDFWCMNLGTVDVHPICWAINSK- 68  
Db 72 NCDSDI---QVRWFARIEKVCGRVLAQFIG---ADTKFWLNLSDDMFGLANAAMSDPN 124  
Qy 69 ----ILVPPRTIHAKE-TWKGYLMK----RLVGSRTLVDVDFHIKMWESMKYPPFGMRL 119  
Db 125 MDKIVVAPPLAINEEYQNDMVNVCIDGIEVGTSLSPKFDGKALLSKHRFKVGQRL 184  
Qy 120 EVDKSCQVSRTRNAVVDVTVIGRLR--LLYED-----GSDDDD-----FWCHMWSP 163  
Db 185 ELLNYSNSTEIRVARIEQICGRMNVSIITKDPFESLPDADDDRVFSSGSQYWIDEGSF 244  
Qy 164 LIHPVGM-----SRVGHGIMKERRSDMAHPTFRKIYCDV-PYLF 205  
Db 245 FIFPVGFAAVNGYQLNKAKKEYIEHTNKIAQAKNGENPRYDSDVTFDQAKDPIDPMIW 304  
Qy 206 KKVAVYTEGGWFEQMKLEAIDPL--NLGNICVATV--CKVLLDGYLMICVDGPGSTD 260  
Db 305 RKVKV-----GQKPELIDPLAQQFNHLVASILKFKCT--EGYLIVGMDGPDAL 352  
Qy 261 GLDMFCYCHASHAIFPATFCQKNDIELTPPKGYEACTFNWENYLEKTKSKAAPSRLFNMD 320  
Db 353 --DNFPIHINNTFMFVGYABKYNLELVPPDFKGTFRWDEYLEKESAEITLPLDLFK-P 408  
Qy 321 CPNH-----GFKVGMKLEAVDLMPEPLICVATVCKVVRHLLSHFGDWDSEYDQ 369  
Db 409 MPQERLDKFKVLLISKVGLRLEAADMCENQFICPATVKSVHGRLLNVNFDGWDSEFDE 468  
Qy 370 WDCESPDYIPVGCWELTGYOLOPP 394  
Db 469 LYVDVSDHILPIGWCEAHSYVLOPP 493

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:07:27 ; Search time 23 Seconds  
(without alignments)  
1675.070 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGMKEVLNSDAVLPSRVY.....KASSPELPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 376054 seqs, 77831530 residues

Total number of hits satisfying chosen parameters: 376054

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_AA\_New.\*

1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/paa/US06\_NEW\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pep.\*  
7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1111	41.4	891	5	US-09-614-150A-38841
2	1111	41.4	1243	5	US-09-614-150A-27069
3	611	22.8	894	6	US-10-719-993-647
4	510	19.0	538	6	US-10-473-040-693
5	510	19.0	538	6	US-10-473-040-780
6	402	15.0	1477	5	US-09-614-150A-12672
7	362.5	13.5	269	6	US-10-473-040-724
8	359	13.4	246	6	US-10-473-040-550
9	327	12.2	877	5	US-09-614-150A-20184
10	109.5	4.1	1026	7	US-60-490-890-1851
11	104	3.9	6669	1	PCT-US03-21379-80
12	104	3.9	6669	1	PCT-US03-38193-25
13	104	3.9	6669	5	US-10-723-860-25
14	102.5	3.8	779	5	US-09-614-150A-31704
15	100.5	3.7	556	6	US-10-425-114A-67814
16	100.5	3.7	524	6	US-10-425-114A-65213
17	98.5	3.7	600	6	US-10-739-930-7654
18	96.5	3.6	1377	7	US-60-490-890-2535
19	96	3.6	1065	6	US-10-648-512-42
20	95.5	3.6	1231	5	US-09-614-150A-6006
21	94	3.5	510	6	US-10-425-114A-49488
22	94	3.5	898	6	US-10-648-512-24
23	94	3.5	906	6	US-10-648-512-36
24	94	3.5	906	6	US-10-648-512-38
25	94	3.5	970	6	US-10-648-512-34
26	94	3.5	1001	6	US-10-648-512-40

27 94 3.5 1065 6 US-10-648-512-22 Sequence 22, Appl  
28 94 3.5 1065 6 US-10-648-512-32 Sequence 32, Appl  
29 93 3.5 414 5 US-09-614-150A-4752 Sequence 4752, Ap  
30 93 3.5 626 6 US-10-679-063-17795 Sequence 17795, A  
31 92.5 3.4 944 6 US-10-425-114A-55600 Sequence 55600, A  
32 92.5 3.4 1578 7 US-60-495-114-1487 Sequence 1487, Ap  
33 92.5 3.4 1593 1 PCT-US03-30720-855 Sequence 855, App  
34 92.5 3.4 1593 1 PCT-US02-14341-14 Sequence 14, Appl  
35 92 3.4 828 5 US-09-614-150A-36534 Sequence 36534, A  
36 92 3.4 1590 7 US-60-495-114-1488 Sequence 1488, Ap  
37 90.5 3.4 557 7 US-60-500-337-2394 Sequence 2394, Ap  
38 90.5 3.4 580 1 PCT-US03-28227-4205 Sequence 4205, Ap  
39 90.5 3.4 580 1 PCT-US03-28227-4206 Sequence 4206, Ap  
40 90.5 3.4 580 1 PCT-US03-28227-4207 Sequence 4207, Ap  
41 90.5 3.4 580 7 US-60-500-337-2393 Sequence 2393, Ap  
42 90.5 3.4 580 7 US-60-500-337-2396 Sequence 2396, Ap  
43 90 3.4 665 5 US-09-614-150A-39579 Sequence 39579, A  
44 90 3.4 1000 6 US-10-275-762-36 Sequence 36, Appl  
45 90 3.4 1026 1 PCT-US03-38193-2539 Sequence 2539, Ap

## ALIGNMENTS

RESULT 1  
US-09-614-150A-38841  
; Sequence 38841, Application US/09614150A  
; GENERAL INFORMATION:  
; APPLICANT: Venter, J. Craig  
; APPLICANT: et al.  
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID  
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE  
; FILE REFERENCE: CLO00728  
; CURRENT APPLICATION NUMBER: US/09/614,150A  
; CURRENT FILING DATE: 2000-07-11  
; PRIOR APPLICATION NUMBER: 60/157,832  
; PRIOR FILING DATE: 1999-10-05  
; PRIOR APPLICATION NUMBER: 60/160,191  
; PRIOR FILING DATE: 1999-10-19  
; PRIOR APPLICATION NUMBER: 60/161,932  
; PRIOR FILING DATE: 1999-10-28  
; PRIOR APPLICATION NUMBER: 60/164,769  
; PRIOR FILING DATE: 1999-11-12  
; PRIOR APPLICATION NUMBER: 60/173,383  
; PRIOR FILING DATE: 1999-12-28  
; PRIOR APPLICATION NUMBER: 60/175,693  
; PRIOR FILING DATE: 2000-01-12  
; PRIOR APPLICATION NUMBER: 60/184,831  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: 60/191,637  
; PRIOR FILING DATE: 2000-03-23  
; NUMBER OF SEQ ID NOS: 43008  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 38841  
; LENGTH: 891  
; TYPE: PRT  
; ORGANISM: DROSOPHILA  
US-09-614-150A-38841

Query Match 41.4%; Score 1111; DB 5; Length 891;

Best Local Similarity 43.4%; Pred. No. 5.8e-75;

Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;

QY 3 GMKVEVLNSD-----AVLPSRVY---YWIASVIOTAGYVLLRYEGFENDASHDFWNCILGT 54  
Db 218 GMKVEVLNSD-----AVLPSRVY---YWIASVIOTAGYVLLRYEGFENDASHDFWNCILGT 276  
QY 55 VDHPHIGKALNSKILVPPRTIHAFTDWKGYLMKRLVSGRTLPVDPHIKRVESMKYPPFR 114  
Db 277 AEVHSVGVKATRGKGLPIPPRTIHAFTDWKGYLMKRLVSGRTLPVDPHIKRVESMKYPPFR 336  
QY 115 QGMRELVVDKRSQVSRTRMAVVDVTIGGRLRLLEDGSDDDDFWCHMWSPLIHPVGMSSRV 174

```

Db 337 LGNLECDVKDRISQVRLATVTKIVGKRLFLRY--FSDDDGFCHEDSPIIHPVGWATTV 394
Qy 175 GHGI-----KMSERRSDMAHPTFRKIYCDVAPYLFK---KYRAVYTEG--GWFEF 220
Db 395 GHNLAAQODYLERMLAGREAMIEVHED-----DATIELFKMFTFDEYSDGKTNSEFVE 448
Qy 221 GMKLEADPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDMFCYHASSHAIFPATF 279
Db 449 GMKLEAVDFNLSSICPATVMVAVLKFGYMMIRIDSYQPDASGSDWFCYHEKSPCIFPAGF 508
Qy 280 CQKNDIELTPKGYEAGTFNENYLEKTKSKAAPSRLENMDCFNHGFVKVGMKLEAVDLME 339
Db 509 CSVNNISVTPNGYDSRTFTWEGYLRDTGAVAAQGLFHRIIPDHGFVGMKLECADLMD 568
Qy 340 PRLICVATVRRVHRLLSIHFDGWDSEYDQWDCESPDYFVGVGCMLTGYQLQPP----- 394
Db 569 PRLCVATVAVRVGRLLKVFHDFGWTDEYDQWDCESADYFVGVGMVLVNHKLEGGPRVAH 628
Qy 395 -VAAEPATPLKAKEATKXKKKQFGKKRIPPTKTRPLRQSGKKPLLEDDPQGARKISSE 453
Db 629 QOAPKPAKPKIORKKRPKKAAGGK-----TPDNNNTQSVKSRITALKTTTTH----- 676
Qy 454 PVPGEIITAVRKEEHLDA-----SPDKASSPELP-----VSVENIKQETDD 495
Db 677 -LPKLSIKLELKEPHHNAAFYENNOPEEGDEEDPDADGDDGSTSHISEQSTTQSSSD 734

```

## RESULT 2

```

US-09-614-150A-27069
; Sequence 27069, Application US/09614150A
; GENERAL INFORMATION:
; APPLICANT: J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/09/614,150A
; CURRENT FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27069
; LENGTH: 1243
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-09-614-150A-27069

```

```

Query Match 41.4%; Score 1111; DB 5; Length 1243;
Best Local Similarity 43.4%; Pred. No. 8.1e-75;
Matches 234; Conservative 75; Mismatches 162; Indels 68; Gaps 14;

```

```

Qy 3 GMKVEVLNSD-----AVLPERSV---YVIASVIQTAGYRVLLRYEGFENDASHDFWNLCT 54
Gb 570 GMRKEVEVTCDSIEVTPQGTPTSFVWATILEIKGYKALMSYEGFDTD--SHDFWNLGN 628
55 VDVHPIGWCAINSKILVPPRTIHAFTDWKGYLMKRLVGSRTLPVDFHIMVSMKYPFR 114

```

```

Db 629 AEVHSVGWCATRGKPLIPPTTIEHKYKDWKDFVGLRSGARTLPSNFYKINDLSQSRFR 688
Qy 115 QGMRELVVDKSOVSRRTMAVVDTVIGGRLLLYEDGSDDDFWCHMWSPLIHPVGMWRRY 174
Db 689 LGNLECDVKDRISQVRLATVTKIVGKRLFLRY--FSDDDGFCHEDSPIIHPVGWATTV 746
Qy 175 GHGI-----KMSERRSDMAHPTFRKIYCDVAPYLFK---KYRAVYTEG--GWFEF 220
Db 747 GHNLAAQODYLERMLAGREAMIEVHED-----DATIELFKMFTFDEYSDGKTNSEFVE 800
Qy 221 GMKLEADPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDMFCYHASSHAIFPATF 279
Db 801 GMKLEAVDFNLSSICPATVMVAVLKFGYMMIRIDSYQPDASGSDWFCYHEKSPCIFPAGF 860
Qy 280 CQKNDIELTPKGYEAGTFNENYLEKTKSKAAPSRLENMDCFNHGFVKVGMKLEAVDLME 339
Db 861 CSVNNISVTPNGYDSRTFTWEGYLRDTGAVAAQGLFHRIIPDHGFVGMKLECADLMD 920
Qy 340 PRLICVATVRRVHRLLSIHFDGWDSEYDQWDCESPDYFVGVGCMLTGYQLQPP----- 394
Db 921 PRLCVATVAVRVGRLLKVFHDFGWTDEYDQWDCESADYFVGVGMVLVNHKLEGGPRVAH 980
Qy 395 -VAAEPATPLKAKEATKXKKKQFGKKRIPPTKTRPLRQSGKKPLLEDDPQGARKISSE 453
Db 981 QOAPKPAKPKIORKKRPKKAAGGK-----TPDNNNTQSVKSRITALKTTTTH----- 1028
Qy 454 PVPGEIITAVRKEEHLDA-----SPDKASSPELP-----VSVENIKQETDD 495
Db 1029 -LPKLSIKLELKEPHHNAAFYENNOPEEGDEEDPDADGDDGSTSHISEQSTTQSSSD 1086

```

## RESULT 3

```

US-10-719-993-647
; Sequence 647, Application US/10719993
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001496
; CURRENT APPLICATION NUMBER: US/10/719,993
; CURRENT FILING DATE: 2003-11-24
; NUMBER OF SEQ ID NOS: 55342
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 647
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-993-647

```

```

Query Match 22.8%; Score 611; DB 6; Length 894;
Best Local Similarity 34.1%; Pred. No. 1.1e-37;
Matches 156; Conservative 65; Mismatches 170; Indels 66; Gaps 15;
Qy 3 GMKVEVLNSDAVLPERSVYVIASVIQTAGYRVLLRYEGFENDASHDFWNLCTVDVHPIG 62
Db 77 GMKLEVANKN---NPDTYVWATILITTCQULLRLYCYGGERRADFWCDVVIADLHPVGW 133
Qy 63 CAINSKILVPPRTIHAFTDWKGYLMKRLVGSRTLPVDFHIMVSMKYPFR----- 114
Db 134 CTQNNKVLMPDPAIKEKYTDWTEFLIRDLTGSRTAPANL-----LEGPLRGKGPIDL 186
Qy 115 -QGMRELVVDKSOVSRRTMAVVDTVIGGRLLLY---EDGSDDDFWCHMWSPLIHPVGMW 170
Db 187 TVGSLIELQDSQNPFPQYVIVSVIENVGGRLLRYVGLDETSYDQ--WLFYLDVYKRLRVGM 245
Qy 171 SRRVGHGKIMERSRD-----MAHH--PTPRKYICDAVPY-----LFFKVRVAYTTE 214
Db 246 CQE-----NKYRMDPSPSEIYPLKMASEWKCTLEKSLIDAAKFPLPMEVFKDHADLRSH 298
Qy 215 GGMFEKGKLEADPLNLGNICVATVCKVLLDGYLMICVDG--GPSTDGLDMFCYHASSHA 273
Db 299 --FTVGMKLETVNMCEFFYIISPASVTKVFNHFFQVTTIDDLPEPSKSLMLC--HADSGLG 355

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 4, 2004, 09:05:51 ; Search time 20 Seconds  
(without alignments)  
2380.175 Million cell updates/sec

Title: US-10-031-915-36

Perfect score: 2683

Sequence: 1 MKGKVEVLSDAVLSRVY.....KASSPELPVSVENIKQETDD 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: PIR1.\*
- 2: PIR2.\*
- 3: PIR3.\*
- 4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	494.5	18.4	619	2 T23964	hypothetical prote
2	471	17.6	711	2 T12525	hypothetical prote
3	402	15.0	1477	2 T13797	tumor suppressor pr
4	384.5	14.3	390	2 T14794	hypothetical prote
5	116.5	4.3	200	2 F64217	ribosomal protein
6	104	3.9	6669	2 S55024	nebulin, skeletal
7	99.5	3.7	1314	2 T09481	mating type silenc
8	98	3.7	871	2 G86586	DNA topoisomerase
9	98	3.7	871	2 D72038	DNA topoisomerase
10	97.5	3.6	739	2 S18642	polynucleotide ade
11	97.5	3.6	1379	2 T45119	FIM protein [impor
12	97	3.6	689	2 S17875	polynucleotide ade
13	96.5	3.6	740	2 S17925	polynucleotide ade
14	95.5	3.6	721	2 A33319	microtubule-associ
15	95	3.5	322	2 B43427	neurofilament prot
16	94	3.5	756	2 T49475	related to tol pro
17	94	3.5	1644	2 F91286	hypothetical prote
18	94	3.5	1644	2 B86128	hypothetical prote
19	93	3.5	701	2 F84912	hypothetical prote
20	93	3.5	1113	2 T47381	hypothetical prote
21	92.5	3.4	463	2 T39004	probable histone a
22	92.5	3.4	526	1 T45058	phosphoprotein pho
23	92	3.4	699	2 T09483	Cys-rich protein R
24	92	3.4	815	2 JG0197	myosin-light-chain
25	92	3.4	869	2 H86440	unknown protein [i
26	91.5	3.4	246	2 S26826	histone H1 - maize
27	91.5	3.4	1285	2 H85041	hypothetical prote
28	91.5	3.4	1493	2 F70435	glutamate synthase
29	91	3.4	401	2 S54437	hypothetical prote

hematopoietic line  
hypothetical prote  
hypothetical prote  
protein F180A4.25  
related to gastrin  
zinc finger protei  
hypothetical prote  
histone H1, stress  
neurofilament trip  
cytochrome P450 A1  
probable Na+/H+ an  
neurofilament trip  
membrane klotho pr  
ankyrin 2, neuroa  
ftek homolog - Str  
histone H1, drough

30 91 3.4 486 2 S07633  
31 91 3.4 574 2 T27100  
32 90.5 3.4 803 2 T18738  
33 90.5 3.4 1314 2 H8327  
34 90.5 3.4 1331 2 T49813  
35 90.5 3.4 2150 1 S27802  
36 90.5 3.4 2150 2 T19450  
37 90 3.4 207 2 T07035  
38 90 3.4 1020 1 QFHUH  
39 89.5 3.3 510 2 JS0724  
40 89.5 3.3 1162 2 E84431  
41 89 3.3 806 2 A43427  
42 89 3.3 1012 2 JC5925  
43 89 3.3 3924 2 S37431  
44 88.5 3.3 929 2 T35683  
45 88 3.3 202 2 S65059

RESULT 1  
T23964  
hypothetical protein R06C7.7 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T23964  
R:Gardner, A.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z19825  
A:Accession: T23964  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-619 <WIL>  
A:Cross-references: EMBL:Z71266; PIDN:CAA95838.1; GSPDB:GN00019; CESP:R06C7.7  
A:Experimental source: clone R06C7  
C:Genetics:  
A:Gene: CESP:R06C7.7  
A:Map position: 1  
A:Introns: 121/3; 182/3; 238/2; 270/3; 547/3; 594/2

Query Match 18.4%; Score 494.5; DB 2; Length 619;  
Best Local Similarity 29.7%; Pred. No. 2.1e-33;  
Matches 132; Conservative 81; Mismatches 149; Indels 83; Gaps 18;

QY 10 NSDAVLPSRYVTHASVIQTAGYVLLRYEGENDASHDFWNLGTVDPVHPICWCAINSK- 68  
DB 193 NCDSI---QVRFAIEKVCGYRVLAQFIG---ADTKFWNLISDDMFLANAAASDPN 245  
QY 69 ----ILVPPRTTHAKF-TDWKGYLMK----RLVGSRTLPVDFHIKQVESMKYPPFQGMRL 119  
DB 246 MDKIVYAPPLAINEEYQNDWYNNVNCIDGEIVGQTSLSPKFDEGKALLSKHRFKVGQRL 305  
QY 120 EVDVKSQVSRRTMAVVDVVGRLR--LLYED-----GSDDD-----FCHWMSP 163  
DB 306 ELLYNSNSTEIRVARIQICRRMNVSTIKDFPESLPDADDQRFVSSGSGQYWDGSGF 365  
QY 164 LHPHVGW-----SRRVGHGKNSERSDMAHPTFRKIYCDAY-PYLF 205  
DB 366 FIFPGFRAVNGYQLNAKVEYIEHTNKIAQIKGENPRYSDVDVTFDLAKDPIDPMIW 425  
QY 206 KKVAVYTTGGFEGFMKLEAIDPL--NLGNI CVATV---CKVLDDGYLMICVDGPGSTD 260  
DB 426 RKVKV-----GQKFELIDPLAQQFNHLVASILKFKCT--EGYLVGMDGPDALE 473  
QY 261 GLDWFCYHASSHAIFPATFCQKNDIELTPKGYEAQTENWENYLEKTSKAAPSELFNMD 320  
DB 474 --DSFPFIHNTFMPVGVAEKYNLELVPPDEFKG-TFRWDEYLEKESAETLPDLFK-P 529  
QY 321 CPNH-----GFKVGMKLEAVDLMPEPLICVATVKRVVHRLLSIHFDGMDSEYDQ 369  
DB 530 MFSQERLDKFKVILLISKVGLRLEAADNCENQFICPATVKSVHGELINVNFDGWDSEDFE 589

A;Residues: 1-1477 <WIS>  
A;Cross-references: EMBL:X80839; NID:g530289; PIDN:CAA56811.1; PID:g3421009  
C;Genetics:  
A;Cross-references: FlyBase:FBgn002441  
A;Introns: 210/3; 746/1

Query Match 15.0%; Score 402; DB 2; Length 1477;  
Best Local Similarity 21.9%; Pred.No.4e-25;  
Matches 117; Conservative 83; Mismatches 159; Indels 176; Gaps 16;

QY 3 GMKVEVLNDVLPSSRVVIASIQAGYRLLRYEGFENDASHDFWCNLGTVDVHPIGW 62  
| | | :  
Db 853 GMKLEAIDPE--NCSLFCVCSEIVEGRYLKLSPDGY--SSMYDEFWNADSQDIPPGW 907  
| | | :  
QY 63 CAINSKILVPPTRIHAKFTDWGKYLMLKLVGSSTLP--VDPHIKVWESM--KYPRQGMR 118  
| | | :  
Db 908 CDETARVLQAPKDYNSERFSSWSRLVK--TGGKAAPRALFGHLNNQQMDVRNGFAVGWH 965  
| | | :  
QY 119 LEVVDSQSQRTRMAVVVTIVIGRLRLLEDGSDDDFWCHMWSPLIHPYGVMSRRVGHGI 178  
| | | :  
Db 966 LEAEDLNDTGKICVATVTDILDERIRVHFDGWDCDYDLMVHIITSPYIHPCGW----- 1017  
| | | :  
QY 179 KGERSDMAHHTPFKIYCDAVPYLFKKRAVITYEGGFEEGKMKEAIDLPLNLGNICVA 238  
| | | :  
Db 1018 -----HEGRQ----- 1022  
| | | :  
QY 239 TVCKVLLDGLYM:CVDGPGSTDCLDFWCFYHASSHAIFPATFCQNKDIELTPPKGEYAQT 298  
| | | :  
Db 1023 -----QLIVPDYQKSAP 1035  
| | | :  
QY 299 NWENYLETKSKAAPERLFNMDCPNHGFKVMKLEAVDLMEPLICVATVKRVVHRLLSI 358  
| | | :  
Db 1036 IWDYIASEVGMGAASKELFTPRQPMF-YORMKLEVVDQBNPLCIIPATVTVTRKGYRVL 1094  
| | | :  
QY 359 HFDGWSDEVDQWDCSPDLYPVGWCELTGYQLQ-PPVAEAPATPLKA----- 405  
| | | :  
Db 1095 HLDCWPTEYYFLUEDSDPDHLPIGWCAATSHSLETTPGYLOPKSVMPCDVEGCGRGNNAK 1154  
| | | :  
QY 406 -----KEATKKKKKOFKKR-KRIPTKTRP--LRQG-----SKKPILLED 442  
| | | :  
Db 1155 RFLNVLHALRECCYPAPENWRQWRSKTVPKRPVAPENIRRGWAKTKRACSEAQAQIKED 1214  
| | | :  
QY 443 DPQARKISSEPVPGEIIIAVRVKEEHLVDVASPKASSP--ELPSVENIKQETDD 495  
| | | :  
Db 1215 SQQ-----BIVYPKVAE----VVQAKRTSPCEEKVVKQKQOKKEED 1253  
| | | :

RESULT 4  
T14794  
hypothetical protein DKFPz586P1522.1 - human  
C;Species: Homo sapiens (man)  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C;Accession: T14794  
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, August 1999  
A;Reference number: Z18180  
A;Accession: T14794  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-390 <KOE>  
A;Cross-references: EMBL:AL110279  
A;Experimental source: adult uterus; clone DKFPz586P1522  
C;Genetics:  
A;Note: DKFPz586P1522.1

Query Match 14.3%; Score 384.5; DB 2; Length 390;  
Best Local Similarity 36.2%; Pred.No.2e-24;  
Matches 94; Conservative 32; Mismatches 95; Indels 39; Gaps 9;

QY 222 MKLEARDPLNLGNICVATVCKVLLDGYLMICVDGGSTDGLDFWCFYHASSHAIFPATFCQ 281  
| | | :  
Db 1 MKLEAVDRMNPSLVCSASVTDV-VDSRFLVHFNDWDT--YDYWC-DPSPFYIHPVGWCQ 56  
| | | :





```
Db 331 VDKTQVSRTRMAVVDVTIGRLRLLEDGSDDDDFWCHWMSPLIHPVGSRRVGHGKM 390
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEEGMKLEAIDPLNLGNICVATV 240
Db 391 SDRCDSMHPPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEEGMKLEAIDPLNLGSI CVAT 450
Qy 241 CKVLLDGYLMICVDGGSTGDLDFWCFYHSHAFPAIFATFCQKNDIELTPPKGYEAQTFFW 300
Db 451 CKVLLDGYLMICVDGGSTGDLDFWCFYHSHAFPAIFATFCQKNDIELTPPKGYETQFFW 510
Qy 301 ENYLEKTSKAAPSRLENMPCNPHGFKVGMKLEAVDLMEPRLICVATVKKVVRHLLSIHF 360
Db 511 EYLEKTSKAAPSRLENMPCNPHGFKVGMKLEAVDLMEPRLICVATVKKVVRHLLSIHF 570
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGKKR 420
Db 571 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGKKR 630
Qy 421 KRIPPTKTRPLROGSKPILDDPOGARKISSEPVGEEIIAVRKEHLDVASPDKASSP 480
Db 631 KRIPPTKTRPLROGSKPILDDPOGARKISSEPVGEEIIAVRKEHLDVASPDKASSP 688
Qy 481 ELPSVENIKQETDD 495
Db 689 QLPLPIESIKQERN 703

RESULT 2
Q8BZ12 PRELIMINARY; PRT; 464 AA.
ID Q8BZ12 AC Q8BZ12;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Similar to hypothetical protein FL20055.
GN AA08199.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC020018; AA020018.1; -.
DR MGD; MGI:2143977; AA08199.
DR InterPro; IPR004092; Mdt.
DR Pfam; PF02820; mdt; 4.
DR SMART; SM00561; Mdt; 4.
KW Hypothetical protein.
SQ SEQUENCE 464 AA; 52721 MW; 31E5C041592BE874 CRC64;

Query Match 52.9%; Score 1418; DB 11; Length 464;
Best Local Similarity 56.4%; Pred. No. 11e-126;
Matches 264; Conservative 67; Mismatches 121; Indels 16; Gaps 4;

Qy 2 KGMKVEVLSDAVLPSPRVWIASVIQTAGYRVLLRYEGENDASHDFWNLGTVDVHPIG 61
Db 10 ENVRIEVPNTDCSLPTKVFVIAGIILAGYNALLRYEGENDSSLDFFWNCIGSDIHPVG 69
Qy 62 WCAINSILVPPRTIHAFTDMKGYLMKLVGSRTLPVDPHIKWVESMKYPFRQGRLEV 121
Db 70 WCAASGKPLVPPRTVQHKYTNKAFVLKRLTGAKTLPPDFSQKVSQVSEMQYFFPKCMRVEV 129
Qy 122 VDKSQVSRTRMAVVDVTIGRLRLLEDG-DSDDDFWCHWMSPLIHPVGSRRVGHGKM 180
Db 130 VDKRHLCTRVAVESVIGRLRLVYEESEDRDFFWCHMSPLIHHIGWSRSGHGRFR 189
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEEGMKLEAIDPLNLGNICVATV 240
Db 190 SDITKQDGH-----FDTPLPAKVEVDOSGEWFFEGMKLEAIDPLNLSTICVATI 242
Qy 122 VDKSQVSRTRMAVVDVTIGRLRLLEDG-DSDDDFWCHWMSPLIHPVGSRRVGHGKM 180
Db 130 VDKRHLCTRVAVESVIGRLRLVYEESEDRDFFWCHMSPLIHHIGWSRSGHGRFR 189
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEEGMKLEAIDPLNLGNICVATV 240
Db 190 SDITKQDGH-----FDTPLPAKVEVDOSGEWFFEGMKLEAIDPLNLSTICVATI 242
Qy 241 CKVLLDGYLMICVDGGSTGDLDFWCFYHSHAFPAIFATFCQKNDIELTPPKGYEAQTFFW 300
Db 243 RKVLADGFLMIGDGEAADGSDWFCVHATSPSIFPVGFCEINMIELTTPRGYTKLPFKW 302
Qy 301 ENYLEKTSKAAPSRLENMPCNPHGFKVGMKLEAVDLMEPRLICVATVKKVVRHLLSIHF 360
Db 303 FDLRETGSIAPVVKFNKDVFNHGRVGMKLEAVDLMEPRLICVATVTRIHLRLRIHF 362
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGKKR 420
Db 363 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGKKR 463

Query Match 52.1%; Score 1398; DB 11; Length 411;
Best Local Similarity 52.9%; Pred. No. 7.5e-125;
Matches 249; Conservative 56; Mismatches 83; Indels 8; Gaps 2;

Qy 2 KGMKVEVLSDAVLPSPRVWIASVIQTAGYRVLLRYEGENDASHDFWNLGTVDVHPIG 61
Db 10 ENVRIEVPNTDCSLPTKVFVIAGIILAGYNALLRYEGENDSSLDFFWNCIGSDIHPVG 69
Qy 62 WCAINSILVPPRTIHAFTDMKGYLMKLVGSRTLPVDPHIKWVESMKYPFRQGRLEV 121
Db 70 WCAASGKPLVPPRTVQHKYTNKAFVLKRLTGAKTLPPDFSQKVSQVSEMQYFFPKCMRVEV 129
Qy 122 VDKSQVSRTRMAVVDVTIGRLRLLEDG-DSDDDFWCHWMSPLIHPVGSRRVGHGKM 180
Db 130 VDKRHLCTRVAVESVIGRLRLVYEESEDRDFFWCHMSPLIHHIGWSRSGHGRFR 189
Qy 181 SERSDMAHPTFRKIYCDVAVPYLFKKVRAVYTEGGWFEEGMKLEAIDPLNLGNICVATV 240
Db 190 SDITKQDGH-----FDTPLPAKVEVDOSGEWFFEGMKLEAIDPLNLSTICVATI 242
Qy 241 CKVLLDGYLMICVDGGSTGDLDFWCFYHSHAFPAIFATFCQKNDIELTPPKGYEAQTFFW 300
Db 243 RKVLADGFLMIGDGEAADGSDWFCVHATSPSIFPVGFCEINMIELTTPRGYTKLPFKW 302
Qy 301 ENYLEKTSKAAPSRLENMPCNPHGFKVGMKLEAVDLMEPRLICVATVKKVVRHLLSIHF 360
Db 303 FDLRETGSIAPVVKFNKDVFNHGRVGMKLEAVDLMEPRLICVATVTRIHLRLRIHF 362
Qy 361 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGKKR 420
Db 363 DGWSEYDQWDCESPDIYPVGCWELTGYQLQPPVAAEPATPLKAKEATKKKKQFGKKR 463
```

RESULT 4	Q9NXXU1	PRELIMINARY;	PRT;	410 AA.	OC Ephydroidea; Drosophilidae; Drosophila.
ID	Q9NXXU1				RN NCBI_TaxID=7227;
AC	Q9NXXU1				RC SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS).
DT	01-OCT-2000 (TrEMBLrel. 15, Created)				RX MEDLINE=20196006; PubMed=10731132;
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)				RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)				RA Anatolides P.G., Scherer S.E., Richards S., Ashburner M., Henderson S.N.,
DE	Hypothetical protein FLJ20055.				RA Sutton G.G., Lewin J.R., Vandell M.D., Zhang Q., Chen L.X.,
OS	Homo sapiens (Human).				RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				RA April J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,
OX	NCBI_TaxID=9606;				RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari L., Beasley E.M.,
RN					RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RP	SEQUENCE FROM N.A.				RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RC	Tissue=Colon;				RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA	Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,				RA de Pablo B., Delcher A., Beng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Ohtani R., Ota T., Suzuki Y., Ohashi M., Nishi T., Shibahara T.,				RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Tanaka T., Nakamura Y., Isegai T., Sugano S.,				RA Foster C., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RT	"NEDO human cDNA sequencing project";				RA Glöck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RL	Submitted (FEB-2000) to the ENBL/GenBank/DBJ databases.				RA Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
DR	EMBL, AK000082; BAA0919.1; ..				RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
DR	InterPro; IPR004092; Mdt.				RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
DR	Pfam; PF02820; mdt; 4.				RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
DR	SMART; SM00561; MDT; 4.				RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
KW	Hypothetical protein.				RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
SQ	SEQUENCE 410 AA; 46717 MW; 659C09639AC02CF3 CRC64;				RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
					RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
					RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
					RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
					RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
					RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
					RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
					RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
					RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
					RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
					RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhou X., Zheng L.,
					RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.C.,
					RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
					RT "The genome sequence of Drosophila melanogaster.";
					RL Science 287:2185-2195(2000).
					CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS. A LONG FORM (SHOWN HERE) AND A
					CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
					DR EMBL; AE003639; AAF53249.1; ..
					DR FlyBase; FBgn0032475; CG16975.
					DR InterPro; IPR004092; Mdt.
					DR InterPro; IPR001660; SAM.
					DR Pfam; PF02820; mdt; 4.
					DR Pfam; PF00536; SAM; 1.
					DR SMART; SM00561; MDT; 4.
					DR SMART; SM00454; SAM; 1.
					DR PROSITE; PS0105; SAM DOMAIN; 1.
					DR Alternative splicing; Hypothetical protein.
					FT VARSPLOC 1 352 MISSING (IN SHORT ISOFORM).
					SQ SEQUENCE 1243 AA; 136036 MW; 35FC45F6298F5BAA CRC64;
Query Match	52.0%;	Score 1395;	DB 4;	Length 410;	
Best Local Similarity	62.9%;	Pred. No. 1.4e-124;			
Matches 249;	Conservative 55;	Mismatches 84;	Indels 8;	Gaps 2;	
QY	2	KGMKEVLNSDAVPSRVYVIATAGYVLLRYEGFENDASHDFWNCNLGVDPVPIG 61			
DB	10	ENVEVEVNTDCSPTKVFVIAIGLVKAGTALLRYEGFENDGGLDFWNCNLGVDPVPIG 69			
QY	62	WCAINSILVPPRTIHAFTDWMKMLKLVGKNTLPVDFPHIKVSMKYPFRQGRLEV 121			
DB	70	WCAASGKPLVPPRTIHAFTDWMKMLKLVGKNTLPVDFPHIKVSMKYPFRQGRLEV 129			
QY	122	VDSQVSTRMVAVDTVIGRLRLYEDG-DSDDDFCHWNSPLIHPVGHRSRVHGIM 180			
DB	130	VDRKLCRTNVAVVESVIGRLRLYEDG-DSDDDFCHWNSPLIHPVGHRSRVHGIM 189			
QY	181	SERSDMAHPTFRKIYCDAPVLFKRVAVYTEGGWFEGMKLEADPLNLGNCVATV 240			
DB	190	SDTFKKQDGH-----FDTPPHLFAKVEVDQSGWFKEGMKLEADPLNLGNCVATV 242			
QY	241	CKVLIDGLMVCVDPGSPSTGLDWFCHVASHAIFPATPCQKNDIELTPPKGYEATFNW 300			
DB	243	RKVLADGLMIGIDGSAADGSDWFCYHATSPISFVPGCEINNIETPPRGYTKLPFKW 302			
QY	301	ENYLEKTSKAAPSRPLNMDCPNHGFKVGMKLEAVDLMEPRLLCVATVVRVHRLLSIF 360			
DB	303	FDYLRTGTSIAAPVKLFKQVFNHGRVGMKLEAVDLMEPRLLCVATVTRIHLRLRIF 362			
QY	361	DGMDSEVDQWDCESPDYVGVCELTGYQLQPPVA 396			
DB	363	DGMEEEYDQWDCESPDYVGVCELTGYQLQPPAS 398			
RESULT 5	Q9VK33	PRELIMINARY;	PRT;	1243 AA.	OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
ID	Q9VK33				OC Neoptera; Endopterygota; Diptera; Brachycera; Muscormora;
AC	Q9VK33;				
DT	01-MAY-2000 (TrEMBLrel. 13, Created)				
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)				
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)				
DE	CG16975 protein.				
GN	CG16975.				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscormora;				
Query Match	41.4%;	Score 1111;	DB 5;	Length 1243;	
Best Local Similarity	43.4%;	Pred. No. 9.7e-97;			
Matches 234;	Conservative 75;	Mismatches 162;	Indels 68;	Gaps 14;	
QY	3	GMKVEVLNSD-----AVLPSRV---YVIASVIOTAGYVLLRYEGFENDASHDFWNCNLGT 54			
DB	570	GKMEVEVNTDCDSIEVLPQGTPTFSFWAILLEIKYKALMSVEGFTD-SHFVWNLGN 628			
QY	55	VDVHPICGACINSKILVPPRTIHAFTDWMKMLKLVGKNTLPVDFPHIKVSMKYPFR 114			
DB	629	AEVHSVGVKATRGKPLIPPTIIEHKYKMDKDFLVGLSGARTLPSNPNKINDSLQSRFR 688			
QY	115	QGMLEVVDSQVSRTEMAVVDVIGRLRLYEDGSDDFWNCNLGVDPVPIG 174			

```
Db 689 LGLNLECVDRISQVRLATVTKVGRFLRY--FSDDDGFWCHEDSPIHFVGMATTV 746
Qy 175 GHGI-----KMSERRSDMAHPTFRKIYCDVAPYLFK--KVRVYVTEG--GWFER 220
Db 747 GHNLAAQDYLRLMLAGREAMIEVHD-----DATIELFKMFTFDEYYSDGKINSFVE 800
Qy 221 GMLKLEADPLNLGNICVATVCKVLLDGYLMICVDG-GPSTDGLDWCFCYHASSHAIFPATF 279
Db 801 GMLKLEADPLNLGICPATVNAVLFYGMIRIDSYQFDASGDWFCYHEKSPCIFPAGF 860
Qy 280 COKNDIELTPPKGYEATFNWENYLEKTKSAAPSRFLFNMDCPNHGFKGMKLEAVDLME 339
Db 861 CSVANI SVTPNGVDSRTFTWEGYLRDTGAVAGCHLFHRIIPDHGFVGNLSLECADLMD 920
Qy 340 PRLICVATVKKVVRHLLSIHFHDGWDSEYDQVDCESPDIYPVGMCELTYGLOQPP----- 394
Db 921 PRLVCVATVARVGRLLKLVHFDGWTDEYDQVDCESADIYPVGMVCLVNHKLGPPRVAH 980
Qy 395 -VAEPATPLKAKEATKKK-KKQFGKKRIPPTKTRPLRQSGKKPLEDDPQGARKISSE 453
Db 981 QOAPKPAKPKIQKRKPKGAAGK-----TPDNTQSVKSRITIAKTPH----- 1028
Qy 454 PVPGEIIAVRVKBEHLDA-----SPDKASSPELP-----VSVENIKQETDD 495
Db 1029 -LPKLSIKLEKPEHNAAFYENNQPEEGDEEDPDADGDGSGTSHISEQSTTQSSD 1086

RESULT 6
ID Q8C1B4 PRELIMINARY; PRT; 359 AA.
AC Q8C1B4;
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE weakly similar to H-L (fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Skin;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RL Nature 420:563-573 (2002).";
DR EMBL; AK028503; BAC25982.1; -.
FT NON TER
SQ SEQUENCE 359 AA; 41035 MW; CA233E1B6B60CA04 CRC64;

Query Match 40.2%; Score 1079; DB 11; Length 359;
Best Local Similarity 56.2%; Pred. No. 1.9e-94;
Matches 209; Conservative 45; Mismatches 102; Indels 16; Gaps 4;

Qy 98 PVDPHIKVSMKYPFPQGMELVVDKSKQSRTRMAVVDTVIGRLRLAYEDG-DSDDDF 156
Db 1 PPDPSQVSESMQYFPKPCRMVEVDKRLCRTRVAVVESVIGRLRLAYESEDRTDDF 60
Qy 157 WCHMWSPLIHPVGRVGHGKMSERRSDMAHPTFRKIYCDVAPYLFKYPVAVTEGG 216
Db 61 WCHMWSPLIHHGWSRSGHRFKRSDITKKQDGH-----FDPHLLFAKVEVDQSGE 113
Qy 217 WFEGMKLEADPLNLGNICVATVCKVLLDGYLMICVDGSPSTGLDWCFCYHASSHAIFP 276
Db 114 WFEGMKLEADPLNLSTICVATIRKVLADGFLMIGIDGSEADGSDWFCYHATSPSIFP 173
Qy 277 ATFCQKNDIELTPPKGYEATFNWENYLEKTKSAAPSRFLFNMDCPNHGFKGMKLEAVD 336
Db 174 VGFCEINNIELTPRPGYTKLPFKMFDYLRGTGSIAPVFLFNKDVNPHGFRVGMKLEAVD 233
```

```
Qy 337 LMEPLICVATVKKVVRHLLSIHFHDGWDSEYDQVDCESPDIYPVGMCELTYGLOQPPVA 396
Db 234 LMEPLICVATVTRIHRLLRIHFHDGWEEDYDQVDCESPDIYPVGMQCLTYGLOQPPAS 293
Qy 397 AEPATPLKAKEATKK-KKQFGKKRIPPTKTRPLRQSGKKPLEDDPQGARKISSEPV 455
Db 294 OSSRESQASGSKQKAKSQYKGHKWTTSQ-----KEELLGDGSDYSLFHGASQOE 346
Qy 456 PGEIIAVRVKSE 467
Db 347 SNGSATVVIKOE 358

RESULT 7
ID Q96C73 PRELIMINARY; PRT; 866 AA.
AC Q96C73;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Similar to RUI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014614; AAH14614.1; -.
DR InterPro; IPR004092; Mbt.
DR InterPro; IPR001660; SMT.
DR Pfam; PF02820; mbt; 4.
DR Pfam; PF00536; SAW; 1.
DR SMART; SM00561; Mbt; 4.
DR SMART; SM00454; SAW; 1.
SQ SEQUENCE 866 AA; 98141 MW; DCE67BF35C413EB7 CRC64;

Query Match 23.1%; Score 620.5; DB 4; Length 866;
Best Local Similarity 31.9%; Pred. No. 4.3e-50;
Matches 161; Conservative 91; Mismatches 205; Indels 47; Gaps 16;

Qy 3 GKMVEV-LNSDAVLPSRVYTIASVIQTAGYRVLLRYEGFENDASHDFWCNLTVDVHPIG 61
Db 53 GMLKLEAVRTD---PETYVATVITTCEQQLLLRYDGYGDRADRWCDIRKADLYPIG 108
Qy 62 WCANSKILVPPRTIHAFTDKGKLMKRLVGSRTLPVDPHIKVSMKYP---FROGMR 118
Db 109 WCEQNKKTLEAPGIRDKVSDWDEFRLQTLTGACSPVPVL-LEGLRGRNPLDLIAFGSR 167
Qy 119 LEVVDKSOVSRTRMAVVDTVIGRLRLIYEDGSDDDDF--WCHMWSPLIHPVGMSSRVGH 176
Db 168 LECQAFQDSLSTWITVTVVENIGRLKLYEGLESSDNYEHWLYYLDPLFHHVGAQQGY 227
Qy 177 GIKMSERRSDMAHPTFRKIYCDVAPYLFK--KVRVYVTEGGWFEGMKLEA 226
Db 228 ELQPPSATRHLKNAEQEILAKVKEEBEESPLSYLFKDKQVIGIHT----FSVNMKLEA 283
Qy 227 IDPLNLGNICVATVCKVLLDGYLMICVDG-GPSTDGLDWCFCYHASSHAIFPATCQKNDI 285
Db 284 VDPNSPGFISATVTVKVFDEKYLFLVEMDDLPENHARRSFVCHADSPGIFPVQMSLKNGL 343
Qy 286 ELTPPKGYEATFNWENYLEKTKSAAPSRFLFNMDCPNHGFKGMKLEAVDLMEPRLICV 345
Db 344 HISPPPGYPSQDFDWDYLYKQCGAAAPQRCFPLPISHEHEFKENMKLEAVNPILPEEVCV 403
Qy 346 ATVKKVVRHLLSIHFHDGWDSEYDQVDCESPDIYPVGMCELTYGLOQPPVAEPATPLK 404
Db 404 ATITAVRGSYLWLGEGSKKPIPECIVSVESMDLPLFGWCETNGHPL-----STPRR 455
Qy 405 AKEATKKKQFGKKRIPPTKTRPLRQSGKKPLEDDPQGARKISSEPVPGIIVAVR 464
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 16:20:44 ; Search time 883 Seconds  
(without alignments)  
10658.756 Million cell updates/sec

Title: US-10-031-915-90  
Perfect score: 2555  
Sequence: 1 agacatccactctatgac.....ttgcttgagaaaaa 2555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2449703 seqs, 1841816367 residues

Total number of hits satisfying chosen parameters: 4899406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	2550.8	99.8	2781	13	US-10-117-722-622
2	2550.8	99.8	2781	15	US-10-037-270-622
3	2538.8	99.4	3188	13	US-10-094-749-917
4	2514.4	98.4	2663	12	US-10-264-237-419
5	1227.8	48.1	2432	12	US-10-108-260A-1658
6	512.2	20.0	513	13	US-10-027-632-70682
7	512.2	20.0	513	13	US-10-027-632-70683
8	512.2	20.0	513	14	US-10-027-632-70682
9	512.2	20.0	513	14	US-10-027-632-70683
10	471.2	18.4	472	13	US-10-027-632-293974
11	471.2	18.4	472	13	US-10-027-632-302414
12	471.2	18.4	472	14	US-10-027-632-293974
13	471.2	18.4	472	14	US-10-027-632-302414
14	438	17.1	492	11	US-09-918-995-22234
15	413.8	16.2	3590	9	US-09-789-919-55

16	411.4	16.1	458	11	US-09-918-995-26775	Sequence 26775, A
17	386	15.1	445	11	US-09-918-995-26098	Sequence 26098, A
c 18	324.4	12.7	476	10	US-09-867-701-701	Sequence 701, App
19	236.8	9.3	581	13	US-10-029-386-142	Sequence 142, App
20	236.8	9.3	2445	13	US-10-027-632-11883	Sequence 11883, A
c 21	236.8	9.3	2445	13	US-10-027-632-11883	Sequence 103678, A
22	236.8	9.3	2445	14	US-10-027-632-11883	Sequence 11883, A
c 23	236.8	9.3	2445	14	US-10-027-632-11883	Sequence 103678, A
24	235.4	9.2	543	13	US-10-029-386-4056	Sequence 4056, App
25	235.4	9.2	558	9	US-09-864-761-9342	Sequence 9342, App
26	235	9.2	238	13	US-10-029-386-13849	Sequence 13849, A
27	232	9.1	232	9	US-09-864-761-25876	Sequence 25876, A
28	232	9.1	232	13	US-10-029-386-17758	Sequence 17758, A
29	135.8	5.3	165	9	US-09-777-564-1597	Sequence 1597, App
30	135.8	5.3	165	15	US-10-015-219-1597	Sequence 1597, App
c 31	121	4.7	283	12	US-10-242-535A-11074	Sequence 11074, A
32	121	4.7	555	9	US-09-864-761-9308	Sequence 9308, App
33	120.8	4.7	557	9	US-09-864-761-9205	Sequence 9205, App
c 34	117.8	4.6	2635	9	US-09-789-404-1	Sequence 1, Appli
35	117	4.6	159	9	US-09-864-761-25857	Sequence 25857, A
36	106.6	4.2	413	11	US-09-918-995-35932	Sequence 35932, A
c 37	104	4.1	138	12	US-10-242-535A-6856	Sequence 6856, App
38	98	3.8	2099	12	US-10-104-047-1542	Sequence 1542, App
39	98	3.8	2845	12	US-10-104-047-292	Sequence 292, App
40	94.6	3.7	174	9	US-09-789-919-8	Sequence 8, Appli
41	92.2	3.6	237	12	US-10-305-720-588	Sequence 588, App
42	91.8	3.6	586	9	US-09-864-761-9319	Sequence 9319, App
43	91.8	3.6	694	13	US-10-027-632-137841	Sequence 137841, A
44	91.8	3.6	694	13	US-10-027-632-137842	Sequence 137842, A
45	91.8	3.6	694	14	US-10-027-632-137841	Sequence 137841, A

ALIGNMENTS

RESULT 1  
US-10-117-722-622  
; Sequence 622, Application US/10117722  
; Publication No. US20030219744A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219744A1 Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2BCIP  
; CURRENT APPLICATION NUMBER: US/10/117,722  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: Pf\_Fl\_genes Version 1.0  
; SEQ ID NO 622  
; LENGTH: 2781  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (263)..(1750)  
US-10-117-722-622

Query Match 99.8%; Score 2550.8; DB 13; Length 2781;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2552; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 AGACATCCACTCTATGACCACTGGAGGATGTGATGAAGGATGAAGGTGGAGTGC 60  
|||||



Qy	2221	GGGAGAGACATTTGGGAGGAGATGGCCTGAGTGTGCATCTTTGGCTCTGTCTACCTGCTCCT	2380
Db	2448	GGGAGAGACATTTGGGAGGAGATGGCCTGAGTGTGCATCTTTGGCTCTGTCTACCTGCTCCT	2507
Qy	2281	GAAGCCCGCGCTAAAAATAAATTCATCCAAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCCTGA	2340
Db	2508	GAAGCCCGCGCTAAAAATAAATTCATCCAAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCCTGA	2567
Qy	2341	CTGGAGGCTCTAGAGAGCTGGGCTTGTATGTTCTTTTGGGCTTTTGTCTCTACTAAATG	2400
Db	2568	TTGGAGGCTCTAGAGAGCTGGGCTTGTATGTTCTTTTGGGCTTTTGTCTCTACTAAATG	2627
Qy	2401	AAGAAACCATGCTCTGGAGGGGCCGTGAACACAGAACCAACCTCAAGACAAAGGATGACAGAGCT	2460
Db	2628	AAGAAACCATGCTCTGGAGGGGCCGTGAACACAGAACCAACCTCAAGACAAAGGATGACAGAGCT	2687
Qy	2461	GGAGGACACATCTAGCTTGCCATTGCACCTTCATCTGGGCTCCCNAGACTCTGTGTGTGAGA	2520
Db	2688	GGAGGACACATCTAGCTTGCCATTGCACCTTCATCTGGGCTCCCNAGACTCTGTGTGTGAGA	2747
Qy	2521	AATTAACCCCGCTCTTGCTTTGAGAAAAA	2554
Db	2748	AATTAACCCCGCTCTTGCTTTGAGAAAAA	2781

## RESULT 2

```

US-10-037-270-622
; Sequence 622, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Felyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 622
; LENGTH: 2781
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (263)..(1750)
; US-10-037-270-622

Query Match          99.8%; Score 2550.8; DB 15; Length 2781;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2552; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
1 AGCAGTCCCACTCTATGACCACTGGGAGATGTGTGAAGAAAGCGATGAACGGTGGAGGTGC 60

```



QY 1141 AGTGGGTGAGCTCGAGTCCCGAGAGATCTATCCCGCTCGGCTGGTGTGAGCTCACCGGCT 1200  
DB 1368 AGTGGGTGAGCTCGAGTCCCGAGAGATCTATCCCGCTCGGCTGGTGTGAGCTCACCGGCT 1427  
QY 1201 ACCAGCTCCAGCTCTCTGTGGCGCGAGAACCGGCGACACCGCTGAAGGCGCAAGAGGCCA 1260  
DB 1428 ACCAGCTCCAGCTCTCTGTGGCGCGAGAACCGGCGACACCGCTGAAGGCGCAAGAGGCCA 1487  
QY 1261 CAAAGAGAAAGAAACAGTTTGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1320  
DB 1488 CAAAGAGAAAGAAAGAAAGTTTGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1547  
QY 1321 GACCCCTCAGACAGGGGTCCAGAGCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380  
DB 1548 GACCCCTCAGACAGGGGTCCAGAGCCCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1607  
QY 1381 AGATCTCGTGGAGCCCTGTTCTCTGGGAGAGATCATTTCTGTGGTGTGAAGGAGAGAGATC 1440  
DB 1608 AGATCTCGTGGAGCCCTGTTCTCTGGGAGAGATCATTTCTGTGGTGTGAAGGAGAGATC 1667  
QY 1441 TAGAGCTGGGCTCGCGGAGAGAGCTTCAAGTCCAGAGTGCCTGTCTCCGTGAGAGAG 1500  
DB 1668 TAGAGCTGGGCTCGCGGAGAGAGCTTCAAGTCCAGAGTGCCTGTCTCCGTGAGAGAG 1727  
QY 1501 TCAAGCAGGAAACAGAGAGCTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGGAAGCC 1560  
DB 1728 TCAAGCAGGAAACAGAGAGCTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGGAAGCC 1787  
QY 1561 AGCCAGCGTTTCTTACACACACACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 1620  
DB 1788 AGCCAGCGTTTCTTACACACACACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 1847  
QY 1621 TCCTCTGTGTAAATCTGCGCGGTGCTGTGAAGGCTGAGCGGTGGAGAGACTGCTGGG 1680  
DB 1848 TCCTCTGTGTAAATCTGCGCGGTGCTGTGAAGGCTGAGCGGTGGAGAGACTGCTGGG 1907  
QY 1681 GTCTCTGGGACCGGCTGTGCTTCTGCTCCCTGCTGGAAGGCTTATATGACGGGC 1740  
DB 1908 GTCTCTGGGACCGGCTGTGCTTCTGCTCCCTGCTGGAAGGCTTATATGACGGGC 1967  
QY 1741 CGCTGAGGCGCCAGAACTGCTGTGTGAACACACTTTTCAGCGAGGTTCACCAAGCTG 1800  
DB 1968 CGCTGAGGCGCCAGAACTGCTGTGTGAACACACTTTTCAGCGAGGTTCACCAAGCTG 2027  
QY 1801 GAAGCTAGTGTGCTGCTTCTTAAGATGGCTCCCGGACCGCGCCAGCGGCTCTAG 1860  
DB 2028 GAAGCTAGTGTGCTGCTTCTTAAGATGGCTCCCGGACCGCGGCTCTAG 2087  
QY 1861 TTGCCAGGATGGGCGCCACACTGTACACTGTGGAATACAGACAGTGAATCTGTCTG 1920  
DB 2088 TTGCCAGGATGGGCGCCACACTGTACACTGTGGAATACAGACAGTGAATCTGTCTG 2147  
QY 1921 CTTGACAGTGTATGTAATAGTTCTAGAGAGCTCTCTGAGCAGGATAGGTCCCT 1980  
DB 2148 CTTGACAGTGTATGTAATAGTTCTAGAGAGCTCTCTGAGCAGGATAGGTCCCT 2207  
QY 1981 GACAGTGTGTGTGTGGGGGAGCTCTGCTCAAAAATTCACCAAGCAGATGCT 2040  
DB 2208 GACAGTGTGTGTGTGGGGGAGCTCTGCTCAAAAATTCACCAAGCAGATGCT 2267  
QY 2041 CTAGAGCTCATGTGTGGTCTCTGCTCTCTAGCTCCCGAGGAGTTGGGAGCCAG 2100  
DB 2268 CTAGAGCTCATGTGTGGTCTCTGCTCTCTAGCTCCCGAGGAGTTGGGAGCCAG 2327  
QY 2101 CTTGTCTCGGAGTGAAGAGAGTGAACAGGATGGATTTGGCGACTGTGTGGTGG 2160  
DB 2328 CTTGTCTCGGAGTGAAGAGAGTGAACAGGATGGATTTGGCGACTGTGTGGTGG 2387  
QY 2161 CTTGAGCTCTTCTGTGTGTGTGAGGAGTGAATCCCTTCTTAAGGAAATGCCCC 2220  
DB 2388 CTTGAGCTCTTCTGTGTGTGTGAGGAGTGAATCCCTTCTTAAGGAAATGCCCC 2447

QY 2221 GGGAGAGCATTTGGAGGAAAGATGGCTGAGTGTGACATTTGGCTCTGTACCTGTCTCT 2280  
DB 2448 GGGAGAGCATTTGGAGGAAAGATGGCTGAGTGTGACATTTGGCTCTGTACCTGTCTCT 2507  
QY 2281 GAAGCCCGCTAAAAATAATTCATCCAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGA 2340  
DB 2508 GAAGCCCGCTAAAAATAATTCATCCAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGA 2567  
QY 2341 CTGAGGCTCTAGAGAGCTGGCTTGTATCTTTTGGCTTTTGTCTTCTTAAATG 2400  
DB 2568 TTGAGGCTCTAGAGAGCTGGCTTGTATCTTTTGGCTTTTGTCTTCTTAAATG 2627  
QY 2401 AAGAAACCATGCTGGAGGGCGGTGAACACAGAACCTCAAGACAGGATGACAGAGCT 2460  
DB 2628 AAGAAACCATGCTGGAGGGCGGTGAACACAGAACCTCAAGACAGGATGACAGAGCT 2687  
QY 2461 GGAGGACACATCTAGCTGCCATTCGCACTTCCAGAGCTCCCGAGAGCTGTGTGTGAGA 2520  
DB 2688 GGAGGACACATCTAGCTGCCATTCGCACTTCCAGAGCTCCCGAGAGCTGTGTGTGAGA 2747  
QY 2521 AATTAAACCCCTGCTTGTCTTGAGAAAAA 2554  
DB 2748 AATTAAACCCCTGCTTGTCTTGAGAAAAA 2781

RESULT 3

US-10-094-749-917  
; Sequence 917, Application US/10094749  
; Publication No. US20030219741A1  
; GENERAL INFORMATION:  
; APPLICANT: ISOGAI, TAKAO  
; APPLICANT: SUGIYAMA, TOMOYASU  
; APPLICANT: OTSUKI, TETSUJI  
; APPLICANT: WAKAMATSU, AI  
; APPLICANT: SATO, HIROYUKI  
; APPLICANT: ISHII, SHIZUKO  
; APPLICANT: YAMAMOTO, JUN-ICHI  
; APPLICANT: ISONO, YUUKO  
; APPLICANT: HIO, YURI  
; APPLICANT: OTSUKA, KAORU  
; APPLICANT: NAGAI, KEIICHI  
; APPLICANT: IRIE, RYOTARO  
; APPLICANT: TAMECHIKA, ICHIRO  
; APPLICANT: SEKI, NAOHICO  
; APPLICANT: YOSHIKAWA, TSUTOMU  
; APPLICANT: OTSUKA, MOTOKUKI  
; APPLICANT: NAGAHARI, KENJI  
; APPLICANT: MASUHO, YASUHIKO  
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA  
; FILE REFERENCE: 084335/0160  
; CURRENT APPLICATION NUMBER: US/10/094,749  
; PRIOR FILING DATE: 2002-03-12  
; PRIOR APPLICATION NUMBER: 60/350,435  
; PRIOR FILING DATE: 2002-01-24  
; PRIOR APPLICATION NUMBER: JP 2001-328381  
; PRIOR FILING DATE: 2001-09-14  
; NUMBER OF SEQ ID NOS: 3381  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 917  
; LENGTH: 3188  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-094-749-917

Query Match 99.4%; Score 2538.8; DB 13; Length 3188;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2540; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AGCAGCTCCACCTCTATGACAGTGGAGAGATGTGATGAAAGGATGAAGGTGAGGTGC 60  
DB 647 AGCAGCTCCACCTCTATGACAGTGGAGAGATGTGATGAAAGGATGAAGGTGAGGTGC 706  
QY 61 TCAACAGTGTGCTGTGCTCCCGAGCGGGGTACTGCGCTCTGTCTATCCAGACAG 120



Db 707 TCAACAGTGATGCTCTGTCTCCAGCGGGTGTAAGTATCGCTCTGTATCCAGACAG 766  
Qy 121 CAGGGTATCGGGTCTGCTTCGGGTATGAAGCTTTGAAATATGACCGCCATGACTTCT 180  
Db 767 CAGGGTATCGGGTCTGCTTCGGGTATGAAGCTTTGAAATATGACCGCCATGACTTCT 826  
Qy 181 GGTGCAACCTGGGAACAGTGTGATGTCCACCCCATTTGGCTGGTGTGCATCAACAGCAAGA 240  
Db 827 GGTGCAACCTGGGAACAGTGTGATGTCCACCCCATTTGGCTGGTGTGCATCAACAGCAAGA 886  
Qy 241 TCCTAGTGGCCCCCAGCAACCATTCATGCCAAGTTTACCGACTGGAAGGGCTACCTCATGA 300  
Db 887 TCCTAGTGGCCCCCAGCAACCATTCATGCCAAGTTTACCGACTGGAAGGGCTACCTCATGA 946  
Qy 301 AACGECTGGTGGGCTCCAGACAGCTTCCCGTGGATTTCCACATCAAGATGTTGGAGACA 360  
Db 947 AACGGCTGGTGGGCTCCAGACAGCTTCCCGTGGATTTCCACATCAAGATGTTGGAGACA 1006  
Qy 361 TGAAGTACCCCTTTAGCAGGGCATCGGCTGGAAGTGTGGACAAGTCCCAGGTGTAC 420  
Db 1007 TGAAGTACCCCTTTAGCAGGGCATCGGCTGGAAGTGTGGACAAGTCCCAGGTGTAC 1066  
Qy 421 GCATCCCATGCTGTGTGCGACACAGTATATCGGGGTGCGCTACCGCTCTCTACGAGG 480  
Db 1067 GCATCCCATGCTGTGTGCGACACAGTATATCGGGGTGCGCTACCGCTCTCTACGAGG 1126  
Qy 481 ATGGTGACAGTACGACGACTTCTGTGTGCGACACATGTGGAGGCCCTGATCCACCCAGTGG 540  
Db 1127 ATGGTGACAGTACGACGACTTCTGTGTGCGACACATGTGGAGGCCCTGATCCACCCAGTGG 1186  
Qy 541 GTTGGTCAGACAGTGTGGGCCACGGCATCAAGATGTCAAGAGGCGAAGTGCATGGCCCC 600  
Db 1187 GTTGGTCAGACAGTGTGGGCCACGGCATCAAGATGTCAAGAGGCGAAGTGCATGGCCCC 1246  
Qy 601 ATCACCCCACTTCCGGGAAGATCTACTGTGATGCCGTTCTTACCTTCTCAAGAGGTTAC 660  
Db 1247 ATCACCCCACTTCCGGGAAGATCTACTGTGATGCCGTTCTTACCTTCTCAAGAGGTTAC 1306  
Qy 661 GACAGTCTACAGAGGCGGTGTTTGAAGAGGGATGAAGCTGGAGGCCATTGACC 720  
Db 1307 GACAGTCTACAGAGGCGGTGTTTGAAGAGGGATGAAGCTGGAGGCCATTGACC 1366  
Qy 721 CCTGAACTTGGGCACATCTCGGTGGCAACTCTGTGAAGGTTCTCTGGAAGTATCC 1426  
Db 1367 CCTGAACTTGGGCACATCTCGGTGGCAACTCTGTGAAGGTTCTCTGGAAGTATCC 1486  
Qy 781 TGATGATCTGTGTGGAAGGCGGCGCTCCACAGATGGCTTGAAGTGTCTGCTACCATG 840  
Db 1427 TGATGATCTGTGTGGAAGGCGGCGCTCCACAGATGGCTTGAAGTGTCTGCTACCATG 1486  
Qy 841 CCTCTTCCACGCGCATCTTCCGCGCCACCTTCTGTGAGAGGAATGACATTTGAGCTCAC 900  
Db 1487 CCTCTTCCACGCGCATCTTCCGCGCCACCTTCTGTGAGAGGAATGACATTTGAGCTCAC 1546  
Qy 901 CGCAAAAGGTTATGAGGCACAGACTTTCATCTGGAGAACTACTTGGAGAGACCAAGT 960  
Db 1547 CGCAAAAGGTTATGAGGCACAGACTTTCATCTGGAGAACTACTTGGAGAGACCAAGT 1606  
Qy 961 CGAAAGCCGCTCATGAGACTTCTTAACTGAATGTCCTTGAAGGCTTCAAGGTGG 1020  
Db 1607 CGAAAGCCGCTCATGAGACTTCTTAACTGAATGTCCTTGAAGGCTTCAAGGTGG 1666  
Qy 1021 GCATGAAGCTGGAGGCGGTGACCTGTATGGAGCCCGGCTCATCTGTGTGGCCAGGTGA 1080  
Db 1667 GCATGAAGCTGGAGGCGGTGACCTGTATGGAGCCCGGCTCATCTGTGTGGCCAGGTGA 1726  
Qy 1081 AACAGTGTGTGATCGGCTCTTCAAGCATCTTGAAGGCTGGGACAGCGAGTACGACC 1140  
Db 1727 AACAGTGTGTGATCGGCTCTTCAAGCATCTTGAAGGCTGGGACAGCGAGTACGACC 1786  
Qy 1141 AGTGGGTGACCTGGAGTCCCAGACATCTACCCGCTCGGCTGTGTGAGTCTACCGGCT 1200

Db 1787 AGTGGGTGACGTGCGAGTCCCCAGACATCTACCCGCTCGGCTGTGTGAGCTACCGGCT 1846  
Qy 1201 ACCAGCTCAGCCTCTCTGTGCGCGAGAACCGGCCACACCCGCTGAAGGCCAAAAGAGCCA 1260  
Db 1847 ACCAGCTCAGCCTCTCTGTGCGCGAGAACCGGCCACACCCGCTGAAGGTCAAAGAGCCA 1906  
Qy 1261 CAAAGCAAGAAAAGAAACAGTTTGGGAGAAAAGAAAAGAAATCCCGCCCACTAAGAGCC 1320  
Db 1907 CAAAGCAAGAAAAGAAACAGTTTGGGAGAAAAGAAAAGAAATCCCGCCCACTAAGAGCC 1966  
Qy 1321 GACCCCTCAGACAGGGGTCCAAGAGGCCCTGTCTGGAGAGCAGCCTCAGGGTGCACAGA 1380  
Db 1967 GACCCCTCAGACAGGGGTCCAAGAGGCCCTGTCTGGAGAGCAGCCTCAGGGTGCACAGA 2026  
Qy 1381 AGATCTCTCGAGGCTGTCTCTGGCGAGATCATTTGCTGTGCTGTGAAGAAAGCATC 1440  
Db 2027 AGATCTCTCGAGGCTGTCTCTGGCGAGATCATTTGCTGTGCTGTGAAGAAAGCATC 2086  
Qy 1441 TAGAGCTGGCCTCGCCCGACAAAGCTTCAAGTCCAGAGCTGCCGTCTCCGTCAGAAACA 1500  
Db 2087 TAGAGCTGGCCTCGCCCGACAAAGCTTCAAGTCCAGAGCTGCCGTCTCCGTCAGAAACA 2146  
Qy 1501 TCAAGCAGGAAAACAGACGACTGAGCTTCTCTGCTCCAGGCTGGCTTCTAGCTGGAAGCC 1560  
Db 2147 TCAAGCAGGAAAACAGACGACTGAGCTTCTCTGCTCCAGGCTGGCTTCTAGCTGGAAGCC 2206  
Qy 1561 AGCCAGAGGTTTCTTACCAACACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 1620  
Db 2207 AGCCAGAGGTTTCTTACCAACACCATGCTCCACCTGACTTTGGCTTGGAGACTGA 2266  
Qy 1621 TCCTCTCTGTGTAATTTCTGCCGCTGTGTGAAGGCTGACGCTGGAGGACCTGTCTGGG 1680  
Db 2267 TCCTCTCTGTGTAATTTCTGCCGCTGTGTGAAGGCTGACGCTGGAGGACCTGTCTGGG 2326  
Qy 1681 GTCTCTCTGGAGCCCGCTGTGTCTGCGCTGCCCTGTGTGGAAGGCTTATATGAAGGCT 1740  
Db 2327 GTCTCTCTGGAGCCCGCTGTGTCTGCGCTGCCCTGTGTGGAAGGCTTATATGAAGGCT 2386  
Qy 1741 CGCTGAGGGCCCCAGAACTCGTCTGTGAACACACTTTCCAGCCAGAGTTTCCCAAGCTG 1800  
Db 2387 CGCTGAGGGCCCCAGAACTCGTCTGTGAACACACTTTCCAGCCAGAGTTTCCCAAGCTG 2446  
Qy 1801 GAACGCTAGCTGCTGTCTTCTTAAAGTGGCTCCCCCGACCCGCTCAG 1860  
Db 2447 GAACGCTAGCTGCTGTCTTCTTAAAGTGGCTCCCCCGACCCGCTCAG 2506  
Qy 1861 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAAGTGTCTGT 1920  
Db 2507 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAAGTGTCTGT 2566  
Qy 1921 CCTGAACGAGTCTATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGGTCCT 1980  
Db 2567 CCTGAACGAGTCTATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGGTCCT 2626  
Qy 1981 GACAGTGAAGTGTGGTGGGGCAGCCTTGCTCAAAAATTCACCAAGCAGAAATGCCT 2040  
Db 2627 GACAGTGAAGTGTGGTGGGGCAGCCTTGCTCAAAAATTCACCAAGCAGAAATGCCT 2686  
Qy 2041 CTGAGCTCATGTGTGCTCTCTGCTCTTCTAGCTCCCCAGGATGTTGGGACCCAG 2100  
Db 2687 CTGAGCTCATGTGTGCTCTCTGCTCTTCTAGCTCCCCAGGATGTTGGGACCCAG 2746  
Qy 2101 CTTGTCTCGGACCTAAGAAAGAGTGAACAGGATGGATTTTGGCGAAGTGTGTGGTGG 2160  
Db 2747 CTTGTCTCGGACCTAAGAAAGAGTGAACAGGATGGATTTTGGCGAAGTGTGTGGTGG 2806  
Qy 2161 CTTGAGCTGCTTCTGTGTGTGAGGACTGACTCCCATTTCTTAAAGAAATGCCCT 2220  
Db 2807 CTTGAGCTGCTTCTGTGTGTGAGGACTGACTCCCATTTCTTAAAGAAATGCCCT 2866  
Qy 2221 GGGAGGACATTTGGAGGAGAGATGGCTGAGTGTGCACTTTGGCTCTGCTACCTGCTCT 2280  
Db 2867 GGGAGGACATTTGGAGGAGAGATGGCTGAGTGTGCACTTTGGCTCTGCTACCTGCTCT 2926



Db 1091 GCATGAAGCTGGAGGCGGTGGACCTGATGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1150  
 QY 1081 AACGAGTGTGATCGGCTCCCTCAGCATCCATTTGACGGCTGGACACGCGATGAGACC 1140  
 Db 1151 AACGAGTGTGATCGGCTCCCTCAGCATCCATTTGACGGCTGGACACGCGATGAGACC 1210  
 QY 1141 AGTGGTGAAGTGGAGTCCCGAGCATCTACCCCGT - CGGCTGTGTGAGCTCAACCGC 1199  
 Db 1211 AGTGGTGAAGTGGAGTCCCGAGCATCTACCCCGTNCGGCTGTGTGAGCTCAACCGC 1270  
 QY 1200 TACCAGCTCCAGCTCTGTGGCGGAGAACCGGCGACACCCGCTGAAGGCGAAGAGGCC 1259  
 Db 1271 TACCAGCTCCAGCTCTGTGGCGGAGAACCGGCGACACCCGCTGAAGGCGAAGAGGCC 1330  
 QY 1260 ACAAGAAGAAAGAAACAGTTTGGGAAAGAAAGAAAGAAATCCCGCCCACTAAGAGC 1319  
 Db 1331 ACAAGAAGAAAGAAACAGTTTGGGAAAGAAAGAAAGAAATCCCGCCCACTAAGAGC 1390  
 QY 1320 CGACCCCTCAGACAGGGGTCCAAAGAGCCCTGTGGAGGAGCAACCTCAGGGTCCAGG 1379  
 Db 1391 CGACCCCTCAGACAGGGGTCCAAAGAGCCCTGTGGAGGAGCAACCTCAGGGTCCAGG 1450  
 QY 1380 AAGATCTGTCCGAGGCTCTCTGCGGAGATCATGTGTCGCTGTGGAAGAGAGCAT 1439  
 Db 1451 AAGATCTGTCCGAGGCTCTCTGCGGAGATCATGTGTCGCTGTGGAAGAGAGCAT 1510  
 QY 1440 CTAGAGCTGGCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTCTCTCGTCCAGAAC 1499  
 Db 1511 CTAGAGCTGGCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTCTCTCGTCCAGAAC 1570  
 QY 1500 ATCAAGCAGAAACAGACAGTGGCTTCCTGCTCCAGCTGGCTCTAGCTGGAGC 1559  
 Db 1571 ATCAAGCAGAAACAGACAGTGGCTTCCTGCTCCAGCTGGCTCTAGCTGGAGC 1630  
 QY 1560 CAGCCAGAGTTTCTTACACACACACACATCCCTCCAGCTGGCTTCCTGCTGGAGCTG 1619  
 Db 1631 CAGCCAGAGTTTCTTACACACACACACATCCCTCCAGCTGGCTTCCTGCTGGAGCTG 1690  
 QY 1620 ATCTCTCTGTGTAATTCGCTCGGTGTGTGAAGGCTGGAGCTGGAGACCTGCTGG 1679  
 Db 1691 ATCTCTCTGTGTAATTCGCTCGGTGTGTGAAGGCTGGAGCTGGAGACCTGCTGG 1750  
 QY 1680 GGTCTCTGGGACCGCTGTTGCTTCTGCTCCCTCCCTGTTGGAAGGCTTATATGACGG 1739  
 Db 1751 GGTCTCTGGGACCGCTGTTGCTTCTGCTCCCTCCCTGTTGGAAGGCTTATATGACGG 1810  
 QY 1740 CCGCTTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCT 1799  
 Db 1811 CCGCTTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCT 1870  
 QY 1800 GGAACGCTAGTCTGTCTTCTTAAGATGCGCTCCCGCCAGCCCGCCAGCGCCCTCA 1859  
 Db 1871 GGAACGCTAGTCTGTCTTCTTAAGATGCGCTCCCGCCAGCCCGCCAGCGCCCTCA 1930  
 QY 1860 GTTGCCAGGATGGGGCCACCACTGTCACTGTGGATACAGACAGTGAACCTGTCT 1919  
 Db 1931 GTTGCCAGGATGGGGCCACCACTGTCACTGTGGATACAGACAGTGAACCTGTCT 1990  
 QY 1920 GCTGAAACGAGTCATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGATAAGTCCCC 1979  
 Db 1991 GCTGAAACGAGTCATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGATAAGTCCCC 2050  
 QY 1980 TGACAGTGAATGTGTGTGGGGGAGCGCTCTGCTTCAAAAATTCAACAGAGATGCC 2039  
 Db 2051 TGACAGTGAATGTGTGTGGGGGAGCGCTCTGCTTCAAAAATTCAACAGAGATGCC 2110  
 QY 2040 TCTCAGCCTCATGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2099  
 Db 2111 TCTCAGCCTCATGTGTGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2170  
 QY 2100 GCTTGTCTGGCAGCTAAGAACAGTACAGGATGTGGATTTTGGCGACCTGTGTGTG 2159  
 Db 2171 GCTTGTCTGGCAGCTAAGAACAGTACAGGATGTGGATTTTGGCGACCTGTGTGTG 2230

QY 2160 GCCTTGAGCTGCTTTCTGTCTTTTGTGAGGACTGACTCCCATTTTCTTAAAGAAATGCCCC 2219  
 Db 2231 GCCTTGAGCTGCTTTCTGTCTTTTGTGAGGACTGACTCCCATTTTCTTAAAGAAATGCCCC 2290  
 QY 2220 CGGGGAGACATTTGGGAGGAGATGGCTCAGTGTGCACCTTTGGCTCTGTACTGTCTCC 2279  
 Db 2291 CGGGGAGACATTTGGGAGGAGATGGCTCAGTGTGCACCTTTGGCTCTGTACTGTCTCC 2350  
 QY 2280 TGAAGCCCGCTTAAATAATTAATTCATCAAGATTCCTTTGTAGTTAAAGGTCAGTTCTG 2339  
 Db 2351 TGAAGCCCGCTTAAATAATTAATTCATCAAGATTCCTTTGTAGTTAAAGGTCAGTTCTG 2410  
 QY 2340 ACTGAGCCCTTAGAGAGCTGGCTGTATGTCTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTT 2399  
 Db 2411 ACTGAGCCCTTAGAGAGCTGGCTGTATGTCTTTGGCTTTTGGCTTTTGGCTTTTGGCTTTT 2470  
 QY 2400 GAAGAAACCATGTCTGGAGGGCGGTGAACACAGAACCCCTCAAGACAAGGATCAGAGC 2459  
 Db 2471 GAAGAAACCATGTCTGGAGGGCGGTGAACACAGAACCCCTCAAGACAAGGATCAGAGC 2530  
 QY 2460 TGGAGGACACATCTAGCTGCCATTGCAACTCACTGGCTCCCGAGACTCTGTGTGAG 2519  
 Db 2531 TGGAGGACACATCTAGCTGCCATTGCAACTCACTGGCTCCCGAGACTCTGTGTGAG 2590  
 QY 2520 AAATTTAAACCCCTGTCTGTCTGAG 2544  
 Db 2591 AAATTTAAACCCCTGTCTGTCTGAG 2615

RESULT 5

US-10-108-260A-1658  
 ; Sequence 1658, Application US/10108260A  
 ; Publication No. US20040005560A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HELIX RESEARCH INSTITUTE  
 ; TITLE OF INVENTION: No. US20040005560A1e1 full length cdna  
 ; FILE REFERENCE: H1-A0106  
 ; CURRENT APPLICATION NUMBER: US/10/108,260A  
 ; CURRENT FILING DATE: 2002-03-27  
 ; NUMBER OF SEQ ID NOS: 5458  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1658  
 ; LENGTH: 2432  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; US-10-108-260A-1658

Query Match 48.1%; Score 1227.8; DB 12; Length 2432;  
 Best Local Similarity 99.4%; Pred. No. 0;  
 Matches 1232; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGCACGTCCACCTCTATGACCTGAGGAGTGTGATGAAGGATGAAGGTGGAGTGC 60  
 Db 719 AGCACGTCCACCTCTATGACCTGAGGAGTGTGATGAAGGATGAAGGTGGAGTGC 778  
 QY 61 TCAACAGTGAATGTGTGCTCCCGAGCGGGTGTACTGTGATCGCTCTGTATCAGACAG 120  
 Db 779 TCAACAGTGAATGTGTGCTCCCGAGCGGGTGTACTGTGATCGCTCTGTATCAGACAG 838  
 QY 121 CAGGGTATCGGTGCTTCCGGTATGAAGGCTTTCAAAATGACGCCAGCATCTTCT 180  
 Db 839 CAGGGTATCGGTGCTTCCGGTATGAAGGCTTTCAAAATGACGCCAGCATCTTCT 898  
 QY 181 GGTGCAACCTGGGAAACAGTGGATGTCCACCCCATTTGGCTGGTGTGCTCAACAGCAAGA 240  
 Db 899 GGTGCAACCTGGGAAACAGTGGATGTCCACCCCATTTGGCTGGTGTGCTCAACAGCAAGA 958  
 QY 241 TCTAGTGTCCCGACCGGACCATTCATGCGCAAGTTCACCGACTGGAAGGCTACCTCATGA 300  
 Db 959 TCTAGTGTCCCGACCGGACCATTCATGCGCAAGTTCACCGACTGGAAGGCTACCTCATGA 1018  
 QY 301 AACGGCTGGTGGGCTCCAGGACGCTTCCCGTGAATTTCCACATCAAGATGTGTGGAGCA 360

```

Db 1019 AACGGCTGGTGGCTCCAGACGCTCCCGTGGATTTCACATCAAGATGGTGGAGACA 1078
Qy 361 TGAAGTACCCCTTTAGGCGAGGCGATCGGCTGGAGTGTGACAGTCCCAAGTGTCTAC 420
Db 1079 TGAAGTACCCCTTTAGGCGAGGCGATCGGCTGGAGTGTGACAGTCCCAAGTGTCTAC 1138
Qy 421 GCATCTCGCATGGCTGTGGTGGACACAGTAAATCGGGGTCGCCTACGGCTCTCTACGAGG 480
Db 1139 GCATCTCGCATGGCTGTGGTGGACACAGTAAATCGGGGTCGCCTACGGCTCTCTACGAGG 1198
Qy 481 ATGGTACAGTACGACGACTTCTGGTGGACACATGTGGAGCCCTGATCCACCCAGTGG 540
Db 1199 ATGGTACAGTACGACGACTTCTGGTGGACACATGTGGAGCCCTGATCCACCCAGTGG 1258
Qy 541 GTTGGTCAAGACCTGTGGGCGACCGCATCAAGATGTCAGAGAGGGGAAAGTGCATGGCCC 600
Db 1259 GTTGGTCAAGACCTGTGGGCGACCGCATCAAGATGTCAGAGAGGGGAAAGTGCATGGCCC 1318
Qy 601 ATCACCACCCACTTCGGAGAGATCTACTGTGATGCGGTTCCCTTACCTCTTCAAGAGGTAC 660
Db 1319 ATCACCACCCACTTCGGAGAGATCTACTGTGATGCGGTTCCCTTACCTCTTCAAGAGGTAC 1378
Qy 661 GAGCAGTCTACACAGAGGCGGTTGTTTGAAGAGGATGAAGTGGAGGCCATTGACC 720
Db 1379 GAGCAGTCTACACAGAGGCGGTTGTTTGAAGAGGATGAAGTGGAGGCCATTGACC 1438
Qy 721 CCCTGAATCTGGCAACATCTGGTGGCAACTGTCTGTAAAGTTCCTCTGGATGATACC 780
Db 1439 CCCTGAATCTGGCAACATCTGGTGGCAACTGTCTGTAAAGTTCCTCTGGATGATACC 1498
Qy 781 TGATGATCTGTGGAGCGGCGGCTCCACAGATGGCTTGGACGTTCTGCTACCATG 840
Db 1499 TGATGATCTGTGGAGCGGCGGCTCCACAGATGGCTTGGACGTTCTGCTACCATG 1558
Qy 841 CCTCTTCCACGCACTCTTCCCGGCCACCTTCTGTGAGAGGAATGACATTTGAGCTCACAC 900
Db 1559 CCTCTTCCACGCACTCTTCCCGGCCACCTTCTGTGAGAGGAATGACATTTGAGCTCACAC 1618
Qy 901 CGCAGAGGTTATGAGGACAGACATTTCACTGGGAGAACTACTTGGAGAGACCAAGT 960
Db 1619 CGCAGAGGTTATGAGGACAGACATTTCACTGGGAGAACTACTTGGAGAGACCAAGT 1678
Qy 961 CGAAGCGCTCCATCGAGACTCTTTAAATGATGATGATGATGATGATGATGATGATGATG 1020
Db 1679 CGAAGCGCTCCATCGAGACTCTTTAAATGATGATGATGATGATGATGATGATGATGATG 1738
Qy 1021 GCATGAAGCTGAGGCGCGTGGACCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
Db 1739 GCATGAAGCTGAGGCGCGTGGACCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1798
Qy 1081 AACGAGTGTGATCGGCTCCCTCAGCATCCACTTTGAGCGGCTGGACAGCGAGTACGACC 1140
Db 1799 AACGAGTGTGATCGGCTCCCTCAGCATCCACTTTGAGCGGCTGGACAGCGAGTACGACC 1858
Qy 1141 AGTGGTGGACTGCGAGTCCCGAGACATCTACCCGCTGGGCTGTGTGAGCTTCAAGGCT 1200
Db 1859 AGTGGTGGACTGCGAGTCCCGAGACATCTACCCGCTGGGCTGTGTGAGCTTCAAGGCT 1918
Qy 1201 ACCAGTCCAGCTCTCTGTGGCGGAGACCCGGCCACAC 1239
Db 1919 ACCAGTCCAGCTCTCTGTGGCGGAGTGTGGGCTCTC 1957

```

RESULT 6

```

US-10-027-632-70682
; Sequence 70682, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129

```

```

; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: PAST-SEQ for Windows Version 4.0
; SEQ ID NO 70682
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-70682

```

Query Match

```

Best Local Similarity 20.0%; Score 512.2; DB 13; Length 513;
Matches 511; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1439 TCTAGAGCTGGCTCGCCCGACAGAGCTTCAAGTCCAGAGCTCCCTGTCTCCCTCGAGAA 1498
Db 1 TCTAGAGCTGGCTCGCCCGACAGAGCTTCAAGTCCAGAGCTCCCTGTCTCCCTCGAGAA 60
Qy 1499 CATCAAGCAGGAAACAGACGACTGAGCTTCTTCCCTCCAGCTGCTTCTAGCTGGAAG 1558
Db 61 CATCAAGCAGGAAACAGACGACTGAGCTTCTTCCCTCCAGCTGCTTCTAGCTGGAAG 120
Qy 1559 CCAGCCAGAGGTTTCTTCTACACACACCATCCCTCCACCTGACTTTGGCTTGGAGACT 1618
Db 121 CCAGCCAGAGGTTTCTTCTACACACACCATCCCTCCACCTGACTTTGGCTTGGAGACT 180
Qy 1619 GATCCTCTCTGTAAATTTCTGCCGGTGTGTGAAGGTGGACGCTGGAGGACCTGCTG 1678
Db 181 GATCCTCTCTGTAAATTTCTGCCGGTGTGTGAAGGTGGACGCTGGAGGACCTGCTG 240
Qy 1679 GGGTCTCTGGGACCGCTGTGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 1738
Db 241 GGGTCTCTGGGACCGCTGTGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 300
Qy 1739 GCGGCTTGAGGCGCCCGAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAAGC 1798
Db 301 GCGGCTTGAGGCGCCCGAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAAGC 360
Qy 1799 TGAACGCTAGCTGCTGCTTCTTCTTAAGATGGCTCCCGCCCGCCCGCCCGCCCGCCCTC 1858
Db 361 TGAACGCTAGCTGCTGCTTCTTCTTAAGATGGCTCCCGCCCGCCCGCCCGCCCGCCCTC 420
Qy 1859 AGTTGCCAGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAAGTCTGTGTC 1918
Db 421 AGTTGCCAGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAAGTCTGTGTC 480
Qy 1919 TGCTGAGCAGCTGATGTAATTAAGTTCTAGA 1951
Db 481 TGCTGAGCAGCTGATGTAATTAAGTTCTAGA 513

```

RESULT 7

```

US-10-027-632-70683
; Sequence 70683, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70683
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70683

```

```

Query Match      20.0%; Score 512.2; DB 13; Length 513;
Best Local Similarity 99.6%; Pred. No. 1e-147;
Matches 511; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1439 TCTAGAGCTGGCTCGCCGACAGAGCTTCAAGTCAGAGCTGCTGCTCGGTGAGAA 1498
Db 1 TCTAGAGCTGGCTCGCCGACAGAGCTTCAAGTCAGAGCTGCTGCTCGGTGAGAA 60

QY 1499 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTGCTAGCTGGAG 1558
Db 61 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTGCTAGCTGGAG 120

QY 1559 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTTCCAGCTGCTTGGCTTGGAGACT 1618
Db 121 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTTCCAGCTGCTTGGCTTGGAGACT 180

QY 1619 GATCCTCTCTGTAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1678
Db 181 GATCCTCTCTGTAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

QY 1679 GGGTCTCTGAGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1738
Db 241 GGGTCTCTGAGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 1739 GCGGCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGC 1798
Db 301 GCGGCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGC 360

QY 1799 TGAAGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1858
Db 361 TGAAGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420

QY 1859 AGTTGCGAGGATGGGGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1918
Db 421 AGTTGCGAGGATGGGGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

QY 1919 TGCCTGAACGAGTCATGTAAATTAAGTTCTAGA 1951
Db 481 TGCCTGAACGAGTCATGTAAATTAAGTTCTAGA 513

```

```

RESULT 8
US-10-027-632-70682
; Sequence 70682, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027.632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70682
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-70682

```

```

Query Match      20.0%; Score 512.2; DB 14; Length 513;
Best Local Similarity 99.6%; Pred. No. 1e-147;
Matches 511; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1439 TCTAGAGCTGGCTCGCCGACAGAGCTTCAAGTCAGAGCTGCTGCTCGGTGAGAA 1498
Db 1 TCTAGAGCTGGCTCGCCGACAGAGCTTCAAGTCAGAGCTGCTGCTCGGTGAGAA 60

QY 1499 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTGCTAGCTGGAG 1558
Db 61 CATCAAGCAGAGAAACAGACGACTGAGCTTCTGCTCCAGCTGCTGCTAGCTGGAG 120

QY 1559 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTTCCAGCTGCTTGGCTTGGAGACT 1618
Db 121 CCAGCCAGAGGTTTCTTCTACCAACACCATGCTTCCAGCTGCTTGGCTTGGAGACT 180

QY 1619 GATCCTCTCTGTAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1678
Db 181 GATCCTCTCTGTAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240

QY 1679 GGGTCTCTGAGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1738
Db 241 GGGTCTCTGAGACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300

QY 1739 GCGGCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGC 1798
Db 301 GCGGCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGC 360

QY 1799 TGAAGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1858
Db 361 TGAAGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420

QY 1859 AGTTGCGAGGATGGGGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1918
Db 421 AGTTGCGAGGATGGGGCCACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480

QY 1919 TGCCTGAACGAGTCATGTAAATTAAGTTCTAGA 1951
Db 481 TGCCTGAACGAGTCATGTAAATTAAGTTCTAGA 513

```

```

RESULT 9
US-10-027-632-70683
; Sequence 70683, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome

```

```

; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 70683
; LENGTH: 513
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-70683

```

```

Query Match      20.0%; Score 512.2; DB 14; Length 513;
Best Local Similarity 99.6%; Pred. No. 1e-147; Indels 0; Gaps 0;
Matches 511; Conservative 2; Mismatches 0;

QY 1439 TCTAGACGTGGCTCCGCCGCAAGGCTTCAAGTCAGAGCTGCTTCTCGTTCGAGAA 1498
Db 1 TCTAGACGTGGCTCCGCCGCAAGGCTTCAAGTCAGAGCTGCTTCTCGTTCGAGAA 60

QY 1499 CATCAAGCAGGAAACAGAGACTGAGCTTCTGCTCCAGCTGCTTCTAGCTGGAG 1558
Db 61 CATCAAGCAGGAAACAGAGACTGAGCTTCTGCTCCAGCTGCTTCTAGCTGGAG 120

QY 1559 CCAGCCAGCGTTTCTCTACCAACAGCAGCTTCTGCTCCAGCTGCTTCTAGCTGGAG 1618
Db 121 CCAGCCAGCGTTTCTCTACCAACAGCAGCTTCTGCTCCAGCTGCTTCTAGCTGGAG 180

QY 1619 GATCCTCTCTGTAAATTCGCGCGTCTGTGAAGCTGACGGTGGAGACCTGCTG 1678
Db 181 GATCCTCTCTGTAAATTCGCGCGTCTGTGAAGCTGACGGTGGAGACCTGCTG 240

QY 1679 GGGTCTCTCTGGACCGCGCTGTGCTTCTGCTCCCTGCTGAAAGTCTATATGACGG 1738
Db 241 GGGTCTCTCTGGACCGCGCTGTGCTTCTGCTCCCTGCTGAAAGTCTATATGACGG 300

QY 1739 GCGGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 1798
Db 301 GCGGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 360

QY 1799 TGAAGCGTAGCTGCTGCTCTTCTTAAGTGGCTCCCGCCAGCCGCGCCGCTC 1858
Db 361 TGAAGCGTAGCTGCTGCTCTTCTTAAGTGGCTCCCGCCAGCCGCGCCGCTC 420

QY 1859 AGTTGCCAGGATGGGGCCCAACACTGTACACTGTGGAATACAGACAGTGAATCTGTG 1918
Db 421 AGTTGCCAGGATGGGGCCCAACACTGTACACTGTGGAATACAGACAGTGAATCTGTG 480

QY 1919 TGCCTGAACGAGTCATGTAAATTAAGTTCTAGA 1951
Db 481 TGCCTGAACGAGTCATGTAAATTAAGTTCTAGA 513

```

```

RESULT 10
US-10-027-632-293974
; Sequence 293974, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 293974
; LENGTH: 472
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-293974

```

```

Query Match      18.4%; Score 471.2; DB 13; Length 472;
Best Local Similarity 99.6%; Pred. No. 4.9e-135;
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1480 TGCTGTCTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCTTCTGCTCCCTCCAG 1539
Db 1 TGCTGTCTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCTTCTGCTCCCTCCAG 60

QY 1540 CCGTGGCTTCTAGCTGAAGCCAGCCAGCGTTTCTTACCAACCAACCACTGCTCCACC 1599
Db 61 CCGTGGCTTCTAGCTGAAGCCAGCCAGCGTTTCTTACCAACCAACCACTGCTCCACC 120

QY 1600 TGACTTTGGCTGGAGACTGATCTCTGTGTAAATTCGCGCGTGTGCTGGAAGCTG 1659
Db 121 TGACTTTGGCTGGAGACTGATCTCTGTGTAAATTCGCGCGTGTGCTGGAAGCTG 180

QY 1660 GACGCTGGAGGACCTCTGCTGGGCTCTCTGCGACCCGCTGTTGCTTCTGCTCCCTCTG 1719
Db 181 GACGCTGGAGGACCTCTGCTGGGCTCTCTGCGACCCGCTGTTGCTTCTGCTCCCTCTG 240

QY 1720 GGAAGGCTCTATATGACGGGCGCGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTC 1779
Db 241 GGAAGGCTCTATATGACGGGCGCGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTC 300

QY 1780 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCTTCTTCCCTTAAGATGCGCTCCCC 1839
Db 301 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCTTCTTCCCTTAAGATGCGCTCCCC 360

QY 1840 CCGACCGCCAGCGGCGCTCAGTTGCCAGGATGGGCGCCACCACTGTCACTGTGGAATA 1899
Db 361 CCGACCGCCAGCGGCGCTCAGTTGCCAGGATGGGCGCCACCACTGTCACTGTGGAATA 420

QY 1900 CAAGACAGTGAATCTGCTGCTGCAACGAGTCATGTAAATTAAGTTCTAGA 1951
Db 421 CAAGACAGTGAATCTGCTGCTGCAACGAGTCATGTAAATTAAGTTCTAGA 472

```

```

RESULT 11
US-10-027-632-302414
; Sequence 302414, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632

```



; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/135,135  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/135,135  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 302414  
; LENGTH: 472  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-302414

Query Match 18.4%; Score 471.2; DB 13; Length 472;  
Best Local Similarity 99.6%; Pred. No. 4.9e-135;  
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1480 TGCTGTCTCCGTCGAGAAATCATAGCAGAAACAGAGAGCTGAGCTTCCTGCTCCAG 1539  
DB 1 TGCTGTCTCCGTCGAGAAATCATAGCAGAAACAGAGAGCTGAGCTTCCTGCTCCAG 60  
QY 1540 CCTGGCTTCTAGCTGGAAGCAGCCAGCGCTTCTTACCAACACCAATGCTCCAC 1599  
DB 61 CCTGGCTTCTAGCTGGAAGCAGCCAGCGCTTCTTACCAACACCAATGCTCCAC 120  
QY 1600 TGACTTTGGCTTGGAGACTGATCTCTCTGTGTAATCTGCGCGGTGCTGGAAGGCTG 1659  
DB 121 TGACTTTGGCTTGGAGACTGATCTCTCTGTGTAATCTGCGCGGTGCTGGAAGGCTG 180  
QY 1660 GACGCTGAGGACCTGCTGGGGTCTCTCTGGGACCGCGCTTCTGCTTCCCTCCCTGT 1719  
DB 191 GACGCTGAGGACCTGCTGGGGTCTCTCTGGGACCGCGCTTCTGCTTCCCTCCCTGT 240  
QY 1720 GGAAGGCTATATGACGGCGCGCTGAGCGCCAGAACTCGTCTGTGAACCACTTTTC 1779  
DB 241 GGAAGGCTATATGACGGCGCGCTGAGCGCCAGAACTCGTCTGTGAACCACTTTTC 300  
QY 1780 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCGCTTCTTAAAGATGCGCTCC 1839  
DB 301 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCGCTTCTTAAAGATGCGCTCC 360  
QY 1840 CCGACCCGCGCAGCGCGCTCAGTTGCCAGGATGGGGCCACCACTGTCACTGTGAATA 1899  
DB 361 CCGACCCGCGCAGCGCGCTCAGTTGCCAGGATGGGGCCACCACTGTCACTGTGAATA 420  
QY 1900 CAAGACAGTGAACCTGTCTGCTGCGTGAACGAGTCAATGTAATTAAGTTCTAGA 1951  
DB 421 CAAGACAGTGAACCTGTCTGCTGCGTGAACGAGTCAATGTAATTAAGTTCTAGA 472

RESULT 12  
US-10-027-632-293974  
; Sequence 293974, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/135,135  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 302414  
; LENGTH: 472  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-302414

; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/135,135  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; PRIOR APPLICATION NUMBER: US 60/135,135  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 293974  
; LENGTH: 472  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-293974

Query Match 18.4%; Score 471.2; DB 14; Length 472;  
Best Local Similarity 99.6%; Pred. No. 4.9e-135;  
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1480 TGCTGTCTCCGTCGAGAAATCATAGCAGAAACAGAGAGCTGAGCTTCCTGCTCCAG 1539  
DB 1 TGCTGTCTCCGTCGAGAAATCATAGCAGAAACAGAGAGCTGAGCTTCCTGCTCCAG 60  
QY 1540 CCTGGCTTCTAGCTGGAAGCAGCCAGCGCTTCTTACCAACACCAATGCTCCAC 1599  
DB 61 CCTGGCTTCTAGCTGGAAGCAGCCAGCGCTTCTTACCAACACCAATGCTCCAC 120  
QY 1600 TGACTTTGGCTTGGAGACTGATCTCTCTGTGTAATCTGCGCGGTGCTGGAAGGCTG 1659  
DB 121 TGACTTTGGCTTGGAGACTGATCTCTCTGTGTAATCTGCGCGGTGCTGGAAGGCTG 180  
QY 1660 GACGCTGAGGACCTGCTGGGGTCTCTCTGGGACCGCGCTTCTGCTTCCCTCCCTGT 1719  
DB 181 GACGCTGAGGACCTGCTGGGGTCTCTCTGGGACCGCGCTTCTGCTTCCCTCCCTGT 240  
QY 1720 GGAAGGCTATATGACGGCGCGCTGAGCGCCAGAACTCGTCTGTGAACCACTTTTC 1779  
DB 241 GGAAGGCTATATGACGGCGCGCTGAGCGCCAGAACTCGTCTGTGAACCACTTTTC 300  
QY 1780 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCGCTTCTTAAAGATGCGCTCC 1839  
DB 301 CAGCCAGAGTTCCCAAGCTGGAACGCTAGCTGCGCTTCTTAAAGATGCGCTCC 360  
QY 1840 CCGACCCGCGCAGCGCGCTCAGTTGCCAGGATGGGGCCACCACTGTCACTGTGAATA 1899  
DB 361 CCGACCCGCGCAGCGCGCTCAGTTGCCAGGATGGGGCCACCACTGTCACTGTGAATA 420  
QY 1900 CAAGACAGTGAACCTGTCTGCTGCGTGAACGAGTCAATGTAATTAAGTTCTAGA 1951  
DB 421 CAAGACAGTGAACCTGTCTGCTGCGTGAACGAGTCAATGTAATTAAGTTCTAGA 472

RESULT 13  
US-10-027-632-302414  
; Sequence 302414, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218



PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 302414  
LENGTH: 472  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-302414

Query Match  
Best Local Similarity 18.4%; Score 471.2; DB 14; Length 472;  
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1480 TGCTCTCTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAG 1539  
1 TGCTCTCTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAG 60  
1540 CTGGCTTCTAGTGGAGGACGAGCCAGCGTTTCTTACCAACACCATGCTCCACC 1599  
61 CTGGCTTCTAGTGGAGGACGAGCCAGCGTTTCTTACCAACACCATGCTCCACC 120  
1600 TGACTTTGGCTTGGAGACTGATCTCTGCTGTAATTTCTGCGCGTGTGTGAAGCTG 1659  
121 TGACTTTGGCTTGGAGACTGATCTCTGCTGTAATTTCTGCGCGTGTGTGAAGCTG 180  
1660 GACGCTGGAGGACCTGCTGGGCTCTCTGGGACCGCGCTGTTGCTTCTGCGCTCCCTG 1719  
181 GACGCTGGAGGACCTGCTGGGCTCTCTGGGACCGCGCTGTTGCTTCTGCGCTCCCTG 240  
1720 GGAAGGCTTATGACGGGCGCGCTGAGCCGCGGCTGCTGCTGTGTGAACACCTTTTC 1779  
241 GGAAGGCTTATGACGGGCGCGCTGAGCCGCGGCTGCTGCTGTGTGAACACCTTTTC 300  
1780 CAGCCAGAGTCCCAAGAGCTGAAAGCTGAGCTGCTGCTTCTTCTTAAAGTGGCTCC 1839  
301 CAGCCAGAGTCCCAAGAGCTGAAAGCTGAGCTGCTGCTTCTTAAAGTGGCTCC 360  
1840 CCGACCGGCGGCGCTGAGTGGCCGAGGATGGGGCCACCACTGTCTACCTGTGAATA 1899  
361 CCGACCGGCGGCGCTGAGTGGCCGAGGATGGGGCCACCACTGTCTACCTGTGAATA 420  
1900 CAAGCAGTGAATCTGCTGCTGCTGAAAGCTGATGTAATTAAGTTCTAGA 1951  
421 CAAGCAGTGAATCTGCTGCTGCTGAAAGCTGATGTAATTAAGTTCTAGA 472

RESULT 14  
US-09-918-995-22234  
Sequence 22234, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 22234  
LENGTH: 492  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature

PRIOR FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/167,363  
PRIOR FILING DATE: 1999-11-23  
PRIOR APPLICATION NUMBER: US 60/156,358  
PRIOR FILING DATE: 1999-09-28  
PRIOR APPLICATION NUMBER: US 60/146,002  
PRIOR FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 325720  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 302414  
LENGTH: 472  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-302414

Query Match  
Best Local Similarity 18.4%; Score 471.2; DB 14; Length 472;  
Matches 470; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

1480 TGCTCTCTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAG 1539  
1 TGCTCTCTCCGTCGAGAACATCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAG 60  
1540 CTGGCTTCTAGTGGAGGACGAGCCAGCGTTTCTTACCAACACCATGCTCCACC 1599  
61 CTGGCTTCTAGTGGAGGACGAGCCAGCGTTTCTTACCAACACCATGCTCCACC 120  
1600 TGACTTTGGCTTGGAGACTGATCTCTGCTGTAATTTCTGCGCGTGTGTGAAGCTG 1659  
121 TGACTTTGGCTTGGAGACTGATCTCTGCTGTAATTTCTGCGCGTGTGTGAAGCTG 180  
1660 GACGCTGGAGGACCTGCTGGGCTCTCTGGGACCGCGCTGTTGCTTCTGCGCTCCCTG 1719  
181 GACGCTGGAGGACCTGCTGGGCTCTCTGGGACCGCGCTGTTGCTTCTGCGCTCCCTG 240  
1720 GGAAGGCTTATGACGGGCGCGCTGAGCCGCGGCTGCTGCTGTGTGAACACCTTTTC 1779  
241 GGAAGGCTTATGACGGGCGCGCTGAGCCGCGGCTGCTGCTGTGTGAACACCTTTTC 300  
1780 CAGCCAGAGTCCCAAGAGCTGAAAGCTGAGCTGCTGCTTCTTCTTAAAGTGGCTCC 1839  
301 CAGCCAGAGTCCCAAGAGCTGAAAGCTGAGCTGCTGCTTCTTAAAGTGGCTCC 360  
1840 CCGACCGGCGGCGCTGAGTGGCCGAGGATGGGGCCACCACTGTCTACCTGTGAATA 1899  
361 CCGACCGGCGGCGCTGAGTGGCCGAGGATGGGGCCACCACTGTCTACCTGTGAATA 420  
1900 CAAGCAGTGAATCTGCTGCTGCTGAAAGCTGATGTAATTAAGTTCTAGA 1951  
421 CAAGCAGTGAATCTGCTGCTGCTGAAAGCTGATGTAATTAAGTTCTAGA 472

RESULT 14  
US-09-918-995-22234  
Sequence 22234, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918,995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235,076  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 22234  
LENGTH: 492  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature

LOCATION: (1)...(492)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-22234

Query Match  
Best Local Similarity 17.1%; Score 438; DB 11; Length 492;  
Matches 438; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1947 CTAGCAGAGCTCTCTGAGCAGGATAAGTCCCTCCCTGACAGTGGTGTGTGGGGCAG 2006  
55 CTAGCAGAGCTCTCTGAGCAGGATAAGTCCCTCCCTGACAGTGGTGTGTGGGGCAG 114  
2007 CCTCTGCTCAAAAATTCACCAAGCAGAGTCCCTCTCAGCCTCATGTGTGGTCTCTGC 2066  
115 CCTCTGCTCAAAAATTCACCAAGCAGAGTCCCTCTCAGCCTCATGTGTGGTCTCTGC 174  
2067 TCTCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTCGGAGGAGTGGAGGAGTGC 2126  
175 TCTCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTCGGAGGAGTGGAGGAGTGC 234  
2127 ACCAGGATGGATTTTGGGACCTGTGTGTGGTGGCTTGTGAGTGTCTTGTGTGTGA 2186  
235 ACCAGGATGGATTTTGGGACCTGTGTGTGGTGGCTTGTGAGTGTCTTGTGTGTGA 294  
2187 GGAAGTGGATTTTGGGACCTGTGTGTGGTGGCTTGTGAGTGTCTTGTGTGTGA 2246  
295 GGAAGTGGATTTTGGGACCTGTGTGTGGTGGCTTGTGAGTGTCTTGTGTGTGA 354  
2247 CTGAGTGTGACATTTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2306  
355 CTGAGTGTGACATTTGGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 414  
2307 AGATTTCTTTTGTAGTAAAGGTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2366  
415 AGATTTCTTTTGTAGTAAAGGTCAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 474  
2367 TATGTTCTTTTGGGCTTTT 2384  
475 TATGTTCTTTTGGGCTTTT 492

RESULT 15  
US-09-789-919-55  
Sequence 55, Application US/09789919  
Patent No. US20020064855A1  
GENERAL INFORMATION:  
APPLICANT: Lemischka, Ihor  
TITLE OF INVENTION: GENES THAT REGULATE HEMATOPOIETIC BLOOD FORMING STEM  
FILE REFERENCE: 2275-1-005  
CURRENT APPLICATION NUMBER: US/09/789,919  
CURRENT FILING DATE: 2001-02-21  
NUMBER OF SEQ ID NOS: 96  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 55  
LENGTH: 3590  
TYPE: DNA  
ORGANISM: Mus musculus  
US-09-789-919-55

Query Match  
Best Local Similarity 16.2%; Score 413.8; DB 9; Length 3590;  
Matches 735; Conservative 0; Mismatches 462; Indels 24; Gaps 2;

1 AGCAGTCTCCACTCTATGACAGTGGGAGGATGTGATCAAGGGGATGAAGTGGAGTGC 60  
750 AGCAGTCTCCACTCTATGACAGTGGGAGGATGTGATCAAGGGGATGAAGTGGAGTGC 809  
61 TCAACAGTGTGCTGTGCTCCAGCCGGGTGTGATCGGATCGCTTCTGTCTATCCAGACAG 120  
810 CCAATACAGACTGCAGTCTACCTACCAAAAGTCTTCTGATGCTGGAATTTATAAATTAG 869

```
QY 121 CAGGGTATCGGGTCTGTCGGTATGAAAGCCTTTGAAATGACGCCAGCCATGACTTCT 180
Db |||||
QY 870 CAGGTATTAATGCCCTTTTGAATATGAAGGATTTGAAATGATTTCTCTGACTTCT 929
Db |||||
QY 181 GGTGCAACCTGGGAACAGTGTATGTCACCCCATTTGGCTGGTGGCCATCAACAGCAAGA 240
Db |||||
QY 930 GGTGCAATATATGTGGGTCTGATATTTATCCAGTTGGTTGGTGGCAGCTAGTGGAAAC 989
Db |||||
QY 241 TCCTAGTGGCCGCCAGGACCAATCCATGCCAAGTTTCCGACTGGAGGGCTACTCATGA 300
Db |||||
QY 990 CTCCTGTTCTCTCTAGAACTGTTCAACATAAATATACAACTGGAAAGCTTTCTAGTAA 1049
Db |||||
QY 301 AACGGCTGGTGGGCTCCAGAGCGCTTCCCTGGATTTCCACATCAAGATGGTGGAGACA 360
Db |||||
QY 1050 AAGACTTACTGGTGCCCAAAACACTTCTCTGATTTCTCAGAAAGGTTTCTGAGAGTA 1109
Db |||||
QY 361 TGAAGTACCCCTTTAGCAGGGCATGGGCTGGAAGTGGTGGACAGTCCCAAGTGTCTAC 420
Db |||||
QY 1110 TGCAATATCTTTCAAACCTTGCATGAGAGTAGAAGTAGTTGACAGAGGCATTTATGTC 1169
Db |||||
QY 421 GCACTCGCATGGCTGTGGTGGACACAGTAAATCGGGGTCCGCTACGGCTCCTCTACGAG 480
Db |||||
QY 1170 GAACAAGTAGCAGTGGTGGAAAGTGTAAATGGAGGACGACTACGGCTGGTGTATGAAG 1229
Db |||||
QY 481 ATGGTGACAGTG---ACGACGACTTCTGGTGGCCACATGTGGAGCCCTGATCCACCCAG 537
Db |||||
QY 1230 AGAGTGAAGATGGAACAGACGACTTCTGGTGGCCACATGCACAGCCCTTAATCCACATA 1289
Db |||||
QY 538 TGGGTTGGTCAACAGCTGTGGGCAAGCATCAAGATGTCAAGAGCGCAAGTGCATGG 597
Db |||||
QY 1290 TTGGATGGTCAAGAAGCATAGGCCATCGATTTCAAGAGATCAGATATTACGAAGAAACAG 1349
Db |||||
QY 598 CCCATCACCCCACTTCCGGAAGATCTACTGTGATGCGGCTTCTTACCTTTCAAGAAG 657
Db |||||
QY 1350 ACGGAC-----ATTGATACCTTCCACACTTATTGTGAAG 1388
Db |||||
QY 658 TAGCAGAGTCTACAGAGGCGGTGGTTTGAAGAGGAGTGAAGCTGGAGGCCATTG 717
Db |||||
QY 1389 TAAAGAAGTAGACAGAGTGGAGAATGGTTCAAGAAGGAATGAAATTTGAAGCTATAG 1448
Db |||||
QY 718 ACCCCCTGATCTGGGCAACATCTGGTGGCACTGTCTGTAAGGTTCTCCTGGATGGAT 777
Db |||||
QY 1449 ACCCATTAATCTTTCTACAAATGTGTGCCACCATTAGAAGGTGTGGCTGATGGAT 1508
Db |||||
QY 778 ACTGATGATCTGTGGACGGGGGCCCTCCACAGATGGCTTGGACTGGTTCTGCTACC 837
Db |||||
QY 1509 TCCTGATGATTGGATTGATGGCTCAGAGCAGCAGATGCTGACTGGTTCTGTTATC 1568
Db |||||
QY 838 ATGCTCTTCCACAGGCATCTTCCGGGCCACCTTCTGTGAGAAGATGACATTTGAGCTCA 897
Db |||||
QY 1569 ATGCAACCTCTCCTCCATTTTCCCTGTGGGTTTCTGTGAATTAACATGATAGAAGTGA 1628
Db |||||
QY 898 CACCGCCAAAAGGTTATGAGGCACAGACTTTCAACTGGGAGAACTACTTGGAGAGACCA 957
Db |||||
QY 1629 CTCCACCCAGAGTTACAAAACCTCTTTTAATGGTTTGACTACCTCAGGGAACCG 1688
Db |||||
QY 958 AGTCGAAACCGCTCCATCGAGACTTTTAAACATGGATGCCCCAAACCATGGCTTTCAAG 1017
Db |||||
QY 1689 GCTCCATTGCAGCACCAGTAAACTATTATTAAGGATGTTCCAAACCAAGGATTCGGT 1748
Db |||||
QY 1018 TGGGCATGAAGCTGGAGGCGGTGGACCTGATGAGCCCGGCTCATCTGTGTGGCCACGG 1077
Db |||||
QY 1749 TAGGAATGAATTAGAGCTGTAGATCTCATGAGGCCCGGTTAATATGTGTAGCCACAG 1808
Db |||||
QY 1078 TGAACGAGTGGTGCATCGGCTCTCAGCATCTTTGACGGCTGGGACACGAGTACG 1137
Db |||||
QY 1809 TTACTCGAATTAATCACCATCTTTGAGGATACATTTTGTGGTTGGGAAGAGTATG 1868
Db |||||
QY 1138 ACCAGTGGGTGACTCCGAGTCCGAGACATCTACCCGTGGCTGGTGTGAGCTCACCG 1197
Db |||||
QY 1869 ACCAGTGGGTGAGTGTAGTCCCTGACCTCTATCTGTAGGGTGGTGTGAGTTAACTG 1928
Db |||||
QY 1198 GCTACGAGCTCCAGGCTCCTG 1218
```

Db 1929 GATATCAACTACAGCCTCCAG 1949

Search completed: February 4, 2004, 20:27:14  
Job time : 892 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 13:33:09 ; Search time 5218 Seconds  
(without alignments)  
11900.716 Million cell updates/sec

Title: US-10-031-915-90

Perfect score: 2555

Sequence: 1 agcagctccactctatgac.....ttgcttgagaaaaa 2555

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_htc:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pin:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_nam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_rtd:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1203.6	47.1	3441	11	BC023933
2	1200.4	47.0	3422	11	AK029115
3	1200.4	47.0	3431	11	AK036510
4	960.4	37.6	1201	13	BX446508

C	5	932.8	36.5	1065	13	BX360491
	6	878	34.4	1201	9	AL560108
	7	872	34.1	1080	12	BM909451
	8	794.6	31.1	900	13	BU184978
	9	783	30.6	927	13	BQ927459
	10	773	30.3	893	14	CA454956
	11	743.2	29.1	876	13	BQ217463
	12	739.8	29.0	851	12	BI092956
	13	737.6	28.9	1057	10	BG396658
	14	734.2	28.7	914	10	BG400812
	15	733.4	28.7	855	13	BI919105
	16	720.8	28.2	915	13	EX445999
	17	704.6	27.6	1201	9	AL581617
C	18	696.2	27.2	736	13	BU634412
C	19	692.6	27.1	739	14	CD365541
C	20	691.4	27.1	821	13	BQ573608
C	21	690.6	27.0	733	10	BG685391
C	22	689.8	27.0	721	12	BM666677
C	23	688.4	26.9	695	12	BM666631
C	24	684	26.8	684	14	CB154753
C	25	682.8	26.7	703	14	CA446367
C	26	681.4	26.7	875	13	BX327327
C	27	673.8	26.4	931	10	BG757068
C	28	673.6	26.4	686	12	BM681205
C	29	670.8	26.3	691	13	BU742966
C	30	662	25.9	1089	12	BM479405
C	31	659	25.8	902	13	BQ438013
C	32	657.2	25.7	667	13	BU737070
C	33	643.8	25.2	664	14	CD370874
C	34	641	25.1	900	10	BE729647
C	35	628.8	24.6	762	13	BQ572663
C	36	614.6	24.1	698	10	BE789481
C	37	613.8	24.0	622	10	BF390805
C	38	607	23.8	749	10	BE386405
C	39	604	23.6	605	12	BM726551
C	40	593	23.2	859	14	CA750138
C	41	588	23.0	854	10	BG258301
C	42	575	22.5	693	10	BE280391
C	43	574.6	22.5	2032	11	BC023876
C	44	571.4	22.4	784	14	CA510759
C	45	569.4	22.3	599	12	BM973511

#### ALIGNMENTS

BC023933 3441 bp mRNA linear HTC 23-SEP-2002  
Mus musculus, Similar to hypothetical protein MGC31247, clone  
INAGR:5325131, mRNA.  
ACCESSION BC023933  
VERSION BC023933.1 GI:23271699  
KEYWORDS HTC.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 3441)  
Strausberg.R.  
Direct Submission  
Submitted (05-FEB-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Jeffrey Green M.D.  
cDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),

Gaithersburg, Maryland;  
 Web site: <http://www.nisc.nih.gov/>  
 Contact: nisc\_mgcnhgr1.nih.gov  
 Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
 Blakesley, R.W., Bouffard, G.G., Green, X., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,  
 Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lario, P., Legaspi, R.,  
 Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,  
 McDowell, J., Pearson, K., Stantripop, S., Thomas, P.J., Touchman, J.W.,  
 Turgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,  
 Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 56 Row: 0 Column: 1  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: Hexamer frequency ORF  
 analysis

This clone has the following problem: frame shifted.

#### FEATURES

Location/Qualifiers  
 1. 3441  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5325131"  
 /tissue\_type="Mammary tumor, C3(1)-Tag model. Infiltrating  
 ductal carcinoma. 5 month old virgin mouse."  
 /clone\_lib="NCI CGAP\_Mam6"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 BASE COUNT 817 a 922 c 970 g 732 t  
 ORIGIN

Query Match 47.1%; Score 1203.6; DB 11; Length 3441;  
 Best Local Similarity 86.5%; Pred. No. 66-273;  
 Matches 1353; Conservative 0; Mismatches 204; Indels 7; Gaps 2;  
 1 AGCAGTCCCACTCTATGACAGTGGAGGATGTGATGAAAGGGATGAAGTGGAGGTGC 60  
 649 AACATGTACCTCTATGACAGTGGAGAGCTCATGAGGGGATGAAGTGGAGTGC 708  
 61 TCAACAGTATCTGTGCTCCACCGCGGTGTACTGGATGCTCTGTCTCATTCAGACAG 120  
 709 TCAACAGAGATCTGTGCTCCACCGCGGTGTACTGGATGCTCTCATTCAGACAG 768  
 121 CAGGATATCGGCTGTGCTTCCGATGAGAGCTTTGAAATGACGCCAGCCATGACTTCT 180  
 769 CTGGTACCGGTGTGCTCCGATGAGAGCTTTGAAATGACGCCAGTCTGACTTCT 828  
 181 GGTGCAACTGGGAACAGTGGATGTCCACCCCATTTGGCTGTGTGCTCCATCAACAGCA 240  
 829 GGTGCAACTGGGAACAGTGGATGTCCACCCCATTTGGCTGTGTGCTCCATCAACAGCA 888  
 241 TCCTAGTGTCCCGCCAGGACCATCCATGACAGTGTCCAGTGTGCTGTGCTCCATCA 300  
 889 TCCTGTGACTTCCAGGACCATCCATGACAGTGTGCTGTGCTGTGCTGTGCTGTGCT 948  
 301 AACCGTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 360  
 949 AGCGGTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1008  
 361 TGAAGTACCTTTAGGACAGGCGATCGGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 420  
 1009 TGAAGTACCTTTCCAGAGGCGATGCGCTGTAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1068  
 421 GCATCTGATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 480  
 1069 GTACCCGCGATGCGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 1128  
 481 ATGGTGAAGTGAACAGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 1129 ATGGTGAAGTGAACAGCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1188

QY 541 GTTGGTCAACGAGTGTGGCCACGGCATCAAGATGTGACAGAGCGGAGTGAATGCGCCC 600  
 DB 1189 GTTGGTCCCGCGTGTGGCCACGGCATCAAGATGTGACAGAGCGGATGTGATGTCTC 1248  
 QY 601 ATCAACCCACCTTCCGGAAGATCTACTGTGATCCCGTCTCTTACCTCTTCAAGAGGTAC 660  
 DB 1249 ATCAACCCACCTTCCGGAAGATCTACTGTGATCCCGTCTCTTACCTCTTCAAGAGGTAC 1308  
 QY 661 GAGCAGTCTACACAGAGCGGTGTGTTGAGGAAGGATGAAGTGAAGGCAATGAACTAGAGCCATTGACC 720  
 DB 1309 GCGTGTCTACACAGAGGTGGTGTTCGAGGAAGGATGAACTAGAGCCATTGACC 1368  
 QY 721 CCTTGAATCTGGGCAACATCTGGTGGCAACTGTCTGTAGGTTCTCTCGATGGATACC 780  
 DB 1369 CTCTGAATCTGGGCAAGTATCTGTGTAGCAACCATCTGCAAGGTGCTCTTGGATGGTTACC 1428  
 QY 781 TGATGATCTGTGTGAGCGGGGCGCTTCCACAGATGGCTTGGACTGGTGTCTGTACCATG 840  
 DB 1429 TGATGATCTGTGTGAGTGGGGGCGCTTCCACGGATGGCTTGGTGTCTGTACCATG 1488  
 QY 841 CCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGACAGGAATGACATTGAGCTCACAC 900  
 DB 1489 CCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGACAGGAATGACATTGAGCTCACAC 1548  
 QY 901 CGCCAAAGGTTATGAGGCAACAGCTTTTAACTGGGAGAACTACTTTGGAGAAAGCAAGT 960  
 DB 1549 CCCCAGGAGGTTATGAGACACAGCTTTTCCCTGGGAGACCTACTTTAGAGAAAGCAAGT 1608  
 QY 961 CGAAGCCGCTCCATCGAGACTCTTTAACTGATGATTTGCCAAACCATGGCTTCAAGGTGG 1020  
 DB 1609 CAAAGCTCTCCAGCAAGACTCTTTAACTGATGATTTGCCAAACCATGGCTTCAAGGTGG 1668  
 QY 1021 GCATGAAGTGGAGCGCTGGAGCTCATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080  
 DB 1669 GCATGAAGTGGAGCGCTGGAGCTCATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1728  
 QY 1081 AACGAGTGGTGCATCGGCTCTCAGATCCACTTTTACCGGCTGGGACAGGAGTACGAC 1140  
 DB 1729 AGCGGTGTGTGCATCGGCTCTCAGATCCACTTTGACCGCTGGGACAGGAGTATGATC 1788  
 QY 1141 AGTGGTGAAGTGCAGTCTCCAGACATCTACCCGCTCGGCTGGTGTGAGCTCACCGGCT 1200  
 DB 1789 AGTGGTGAAGTGCAGTCTCCAGACATCTACCCGCTCGGCTGGTGTGAGCTCACCGGCT 1848  
 QY 1201 ACCAGTCTCAGGCTCTGTGTGGCGAGAACCGGCGCACACCGCTGAAGGCCAAAGAGGCA 1260  
 DB 1849 ACCAGTCTCAGGCTCTGTGTGGCGAGAACCGGCGCACACCGCTGAAGGCCAAAGAGGCA 1908  
 QY 1261 CAAAGAGAAAGAAACAGTTTGGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1320  
 DB 1909 CAAAGAGAAAGAAACAGTTTGGGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1968  
 QY 1321 GACCCCTCAGACAGGGTCCAGAGAGCCCTGTGAGAGACGACCCCTCAGGGTGCAGGA 1380  
 DB 1969 GCGCCCTCAGACAGGGTCCAGAGAGCCCTGTGAGAGACGACCCCTCAGGGTGCAGGA 2027  
 QY 1381 AGATCTCGTGGAGCGCTGTTCTCGCGAGATCATTTGCTGTGCTGTGAGGAGAGGATC 1440  
 DB 2028 -----GTCTCGGAAACGAGTCTCTGTGATGACATTTATCGCTGTGTGTGAGGAGGAGCACC 2082  
 QY 1441 TAGAGTGGCCCTCGCGGCAAGGCTTCAAGTCCAGAGTGCCTGTCTCTCGTGCAGAAACA 1500  
 DB 2083 AGGAAATTTCTCTCGGCTGACAGGTGCGCCAGTCCACAGCTGCTCTTCCATTTGAGAGCA 2142  
 QY 1501 TCAAGCAGGAAACAGACGACTGAGGCTT-CCTGGCTCCAGGCTGGCTTCTAGCTGGAAGC 1559  
 DB 2143 TCAGCAGGAGGAGGAACAACTGAGACTTCCCTTGGCATCAGCCTGGAACCCCTAACTGAAGCC 2202  
 QY 1560 CAGC 1563  
 DB 2203 AAGC 2206

## RESULT 2

## AK029115

## LOCUS

## DEFINITION

AK029115 3422 bp mRNA linear HTC 05-DEC-2002  
Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched  
library, clone:4732493N06 product:H-L(3)MBT-LIKE PROTEIN  
(HYPOTHETICAL 79.1 KDA PROTEIN) homolog [Homo sapiens], full insert  
sequence.

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

AK029115.1 GI:26325101

HTC; CAP trapper.

Mus musculus (house mouse)

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

Garninci, P. and Hayashizaki, Y.

High-efficiency full-length cDNA cloning

Meth. Enzymol. 303, 19-44 (1999)

99279253

10349636

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

Garninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,

Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

20493374

11042159

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,

Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,

Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,

Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,

Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M.,

Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,

Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

20530913

11076861

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,

Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,

Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Konno, S., Yamanaka, I.,

Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,

Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,

Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,

Kuehl, P., Lewis, S., Mateu, Y., Nikaide, I., Pesole, G.,

Quackenbush, J., Schriml, L. M., Stauble, F., Suzuki, R., Tomita, M.,

Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,

Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,

Carninci, P., de Bona, M. F., Brownstein, M. J., Bult, C.,

Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,

Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,

Marchionni, L., Mashima, J., Mazza, J., Mombert, P., Nordone, P.,

King, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,

Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. F., Suzuki, H.,

Toyo-oka, K., Wang, K. H., Weitz, C., Whittaker, C., Wilming, L.,

Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S.

and Hayashizaki, Y.

Functional annotation of a full-length mouse cDNA collection

Nature 409 (6821), 685-690 (2001)

21085660

11217851

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

The FANTOM Consortium and the RIKEN Genome Exploration Research

Group Phase I & II Team.

Analysis of the mouse transcriptome based on functional annotation

of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

6 (bases 1 to 3422)

## AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hitamoto, K., Hiraoka, T., Hirozane, T.,  
Hori, P., Imorani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kurahara, C., Kojima, Y., Kondo, S., Konno, H., Koida, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.

## Direct Submission

## TITLE

## JOURNAL

## MEDLINE

## PUBMED

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome-gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

## COMMENT

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues

Please visit our web site for further details.

URL: http://genome-gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

## FEATURES

## source

1. 3422  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM DB:4732493N06"  
/db\_xref="taxon:10090"  
/clone="4732493N06"  
/tissue\_type="skin"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days neonate"

## CDS

58. 2169  
/note="unnamed protein product; H-L(3)MBT-LIKE PROTEIN  
(HYPOTHETICAL 79.1 KDA PROTEIN) homolog [Homo sapiens]  
(SPTRIQ969R5, evidence: FASTY, 90.9%ID, 100%length,  
match=2109)  
putative"

/codon\_start=1

/protein\_id="BAC26305.1"

/db\_xref="GI:26325102"

/translation="MEKPRGTEAPSEPEMEEDDLDFGGYDSFRSYNSAGSES  
SYLSESESENEDEAGELPTSPHLFSSANNRSLDGSSEPAVCEMCGIVGTREAF  
FSKTRFCVSCRSYSSNSKASILARLOQKPPKAKVLAHKAWSAKIGAFHQAQ  
TGQADGTGQDGLVGFDMCKFLKHSYKAPVSCFQGVPLVDQWEDVMKGVY  
LNSDVLPSRVYATVIAQAGYVLLRYEGENDASHDFWNLGTVDPHPIGWCAIN  
SKLVFPRTHAKFTDMKSLMKRLVGSRTLPADPHMKVSMKYPFRQGMLEVVDK  
TQVSRMAVVDITVIGRLLYEDGSDDDFWCHWMSPLIHPVGSRRVGHGKMSD  
RCMDSHPTFRKICDAPVYLFKRVYTEGFWFEQMKLEADPLNLGICVATI  
CKVLLGYLMI CVDGPSDTGDFWCFYHASSHAIFPATFCQKNDILTPKGYETQPF  
AVETLYETKSKAAPARLPNMDCPNHPKVMKLEAVDLMEPRLLICVATVKKVHRL  
SHFDGMDNEYDQWVDCSPDLYPVGWCELTGVLQPPVSAEPNTFOKQDTPKKKK  
QKQKKEKLPKSAKTRPLRGSKKPLLEDNLALGVSEPVDPDIIAIVCKEERQDIISS  
DASPSQLPLPESINQERN"

polyA\_signal

3404. 3409

/note="putative"

polyA\_site

3422

/note="putative"

BASE COUNT 800 a 917 c 966 g 739 t

## ORIGIN

Query Match 47.0%; Score 1200.4; DB 11; Length 3422;

Best Local Similarity 86.4%; Pred. No. 3.4e-272;

Matches 1351; Conservative 0; Mismatches 206; Indels 7; Gaps 2;

## QY

1 AGCACGTCCCACTTATGACCATGGAGGATGTGATGAAGGATGAAGTGGAGGTGC 60

Db 653 AACATGATCCCTCTATGACCAAGTGGGAAGACGTCATGAAGGGGATGAAGTGGAGTGC 712  
 Qy 61 TCAACAGTGTGCTGTGCTCCCGAGCGGGTGTACTGATCCGCTCTGTGATCCAGACAG 120  
 Db 713 TCAACAGCATGTGTGCTCCCGAGCGGGTGTACTGATCCGCTCTGTGATCCAGGCGAG 772  
 Qy 121 CAGGGTATCGGGTGTGCTGTGCTCCGATGAAGGGCTTTGAAATGACGCCAGCCATGACTTCT 180  
 Db 773 CTGGGTACCGGGTGTGCTCCGATGAAGGGCTTTGAAATGACGCCAGCCATGACTTCT 832  
 Qy 181 GGTGCAACCTGGGAACAGTGGATGTCCACCCATTTGGCTGGTGGTGGCTTCAACAGCAAGA 240  
 Db 833 GGTGCAACCTGGGAACAGTGGATGTCCACCCATTTGGCTGGTGGTGGCTTCAACAGCAAGA 892  
 Qy 241 TCTAGTGTCCCGCCACGGAACATCCATGCCCAAGTTCAACCGACTGGAAGGGCTACCTCATGA 300  
 Db 893 TCTGTGTACTCCACGGACCATCCATGCCCAAGTTCAACCGACTGGAAGGGCTACCTCATGA 952  
 Qy 301 AACGGCTGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGFTGGAGACA 360  
 Db 953 AGCGGTGTGGGTTCCAGGACACTTCTCTGCAGACTTCCATATCAAGATGFTGGAGACA 1012  
 Qy 361 TGAAGTACCCCTTTAGGCAAGGCGATCGGCTGGAAGTGTGGACAGTCCAGGTGTAC 420  
 Db 1013 TGAAGTACCCCTTTAGGCAAGGCGATCGGCTGGAAGTGTGGACAGTCCAGGTGTAC 1072  
 Qy 421 GCACTGCGATGCTGTGGTGGACACAGTAAATCGGGGGTGGCTACCGGTCTCTACGAGG 480  
 Db 1073 GTACCGCATGCGCGTGTGGACACAGTAAATCGGGGGTGGCTACCGGTCTCTATGAGG 1132  
 Qy 481 ATGGTGACAGTGCAGCAGACTTCTGGTGCCACATGTGGAGCCCTGTATCCACCGAGTG 540  
 Db 1133 ATGGTGACAGTGCAGCAGACTTCTGGTGCCATGTGGAGTCCCTGTATCCACCGAGTG 1192  
 Qy 541 GTTGTGTACAGCTGTGGGCCCGCATCAAGATGTCAAGAGGCGGAAGTGCATGGGCC 600  
 Db 1193 GTTGTGTCCCGGCTGTGGGCCCGCATCAAGATGTCAAGAGGCGGAAGTGCATGGTCT 1252  
 Qy 601 ATCACCCACCTTCGGAGATCTACTGTAGTGGCTTCTTACCTTCTTCAAGAGGTAC 660  
 Db 1253 ATCACCCACCTTCGGAGAAATCTACTGTAGTGGCTTCTTACCTTCTTCAAGAGGTAC 1312  
 Qy 661 GAGCAGTGTACACAGAGCGGTGTGGTTGAGGAAGGATGAAGTGTGAGGCCATTGACC 720  
 Db 1313 GCGCTGTCTACAGAGAGTGTGGTTCGAGAGAGGATGAAGTGTAGAGGCCATTGACC 1372  
 Qy 721 CCTGAATCTGGCAACATCTCGTGGCAACTGTCTGTAAAGTTCCTCTGGATGGATACC 780  
 Db 1373 CTCTGAATCTGGCAAGTATCTGTAGCAACCATCTGCAAGGTGCTCTTGGATGGTTACC 1432  
 Qy 781 TGATGATCTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGTACCATG 840  
 Db 1433 TGATGATCTGTGAGTGGGGGGCCCTCCACGGATGGCTTCGATGGTCTGTCTACCAAG 1492  
 Qy 841 CTCTTTCACGCCCATCTTCCCGGCCACCTTCTGTCAAGAAGATGACATTGAGTCTCAC 900  
 Db 1493 CTCTTTCACGCCCATCTTCCCGGCCACCTTCTGTCAAGAAGATGACATTGAGTCTCAC 1552  
 Qy 901 CGCCAAAAGTTATGAGGACAGACTTCACTGGAGAACTACTTGGAGAAAGCAAGT 960  
 Db 1553 CCCCAAAAGGTATGAGACACAGCCTTTGGCTGGAGACCTTACTTAGAAGAACCAAGT 1612  
 Qy 961 CGAAAGCCCGCTCCATCGAGACTCTTTAAATGATGGATGGCCCAACCATGGCTTCAAGGTG 1020  
 Db 1613 CAAGAAGCTCTCCAGCAAGACTCTTTAAATGATGGATGGCCCAACCATGGCTTCAAGGTG 1672  
 Qy 1021 GCATGAAGCTGAGGCGGTGGACCTGTATGGAGCCCGCGCTCATCTGTGTGGCCACGGTGA 1080  
 Db 1673 GCATGAAGCTGAGGCGGTGGACCTGTATGGAGCCCGCGCTCATCTGTGTGGCCACCTGTGA 1732  
 Qy 1081 AACGAGTGTGATCGGCTCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140

Db 1733 AGCGGTGTGTCATCGGCTCTCAGCATCCACTTCGACGGCTGGGACAATGATGATGATC 1792  
 Qy 1141 AGTGGTGGACTCGGAGTCCCGAGACATCTACCCCTCGCTGGTGTGAGTCTACCGGCT 1200  
 Db 1793 AGTGGTGGACTCGGAATCCCGGACATCTACCCCTCGCTGGTGTGAGTCTACCGGCT 1852  
 Qy 1201 ACCAGTCTCCAGGCTCTCTGTGGCCGAGAACCGGCCACACCGCTGGAAGGCCAAAGAGGCCA 1260  
 Db 1853 ACCAGTCTCCAGGCTCTCTGTGGCCGAGAACCGGCCACACCTCAGAAAGGCNAGGACACCA 1312  
 Qy 1261 CAAAGAAAGAAAGAAACAGTTTGGGAAGAAAGAAAGAAATCCGCCCATCAAGACGC 1320  
 Db 1913 CAAAGAAAGAAAGAAACAGTTTGGGAAGAAAGAAAGAAATCCCATCAGCCAAAGACTC 1972  
 Qy 1321 GACCCCTCAGACAGGGGTCCAAAGAACCCCTGTCTGAGGACGACCCCTCAGGGTGCAGGA 1380  
 Db 1973 GGCCTCTCAGACAGGGCTCCAAAGAACCCCTTACTTGAGGACCAACCTTGAGGCTTTGGGG- 2031  
 Qy 1381 AGATCTCTCGGAGCTTCTCTGCGGAGATANTCTGTGCGTGTGAAGGAAGAGCATC 1440  
 Db 2032 -----GTCTCGGAACCAAGTTCTCTGTATGATGATTCCTGTGTGTGAAGGAGGAGCACC 2086  
 Qy 1441 TAGAGTGGCTCGCCCGGACCAAGGCTTCAAGTCCAGAGCTGCTGTCTCGTCGGAACA 1500  
 Db 2087 AGGACATTTCTCTGCTGTGAGAGTCCCGCAGTCCACAGCTGCTCTTCCATTTGAGAGA 2146  
 Qy 1501 TCAAGCAGGAAACAGACGACTGAGCCCTT-CCTGCTCCAGCCCTGGCTTCTAGCTGGAAGC 1559  
 Db 2147 TCAAGCAGGAGAGGAACAATGAGACTTCCCTGGCATCAGCTTGGACCCCTTACTGAAGCC 2206  
 Qy 1560 CAGC 1563  
 Db 2207 AAGC 2210

RESULT 3  
 AK036510  
 LOCUS  
 DEFINITION  
 AK036510 3431 bp mRNA linear HTC 05-DEC-2002  
 Mus musculus adult male bone cDNA, RIKEN full-length enriched  
 library, clone:19830123N10 product:H-L(3)MBT-LIKE PROTEIN  
 (HYPOTHEICAL 79.1 KDA PROTEIN) homolog (Homo sapiens), full insert  
 sequence.  
 ACCESSION  
 AK036510  
 VERSION  
 AK036510.1 GI:263331451  
 KEYWORDS  
 HTC; CAP trapper.  
 SOURCE  
 Mus musculus (house mouse)  
 ORGANISM  
 Mus musculus  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 1  
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
 Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayaishizaki, Y.  
 High-efficiency full-length cDNA cloning  
 Meth. Enzymol. 303, 19-44 (1999).  
 2  
 Normalisation and subtraction of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes  
 Genome Res. 10 (10), 1617-1630 (2000)  
 3  
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayaishizaki, Y.  
 RIKEN integrated sequence analysis (RISA) system--384-format  
 sequencing pipeline with 384 multicapillary sequencer

JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

Genome Res. 10 (11), 1757-1771 (2000)  
20530913  
11076861  
4  
Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G., Tomita, M.,  
Quackenbush, J., Schriml, L.M., Staehli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barh, G., Blake, J., Boifelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Rang, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,  
Wyshak-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohtsuki, S.,  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)  
21085660  
11217851  
5

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

The PANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
Nature 420 563-573 (2002)  
6 (bases 1 to 3431)  
Fukuda, S., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,  
Adachi, J., Furuno, M., Haragaki, T., Hara, A., Hashizume, W.,  
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozawa, T.,  
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,  
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,  
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,  
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Tagami, M., Takahashi, F., Takaku-Akahira, S.,  
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,  
Muramatsu, M. and Hayashizaki, Y.  
Direct Submission  
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. 3431  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:9830123N10"  
/db\_xref="taxon:10090"  
/clone="9830123N10"  
/sex="male"  
/tissue\_type="bone"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"

TITLE  
JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. 3431  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:9830123N10"  
/db\_xref="taxon:10090"  
/clone="9830123N10"  
/sex="male"  
/tissue\_type="bone"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"

FEATURES  
source

CDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site for further details.  
URL: http://genome.gsc.riken.go.jp/  
URL: http://fantom.gsc.riken.go.jp/  
Location/Qualifiers  
1. 3431  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="FANTOM:DB:9830123N10"  
/db\_xref="taxon:10090"  
/clone="9830123N10"  
/sex="male"  
/tissue\_type="bone"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"

/dev stage="adult"  
66. -2177  
/note="unlabeled protein product; H-L(3)MPT-LIKE PROTEIN  
(HYPOPHYSICAL 79.1 KDA PROTEIN) homolog [Homo sapiens]  
match=996985, evidence: FASTY, 90.9%ID, 100%length,  
putative"  
/codon\_start=1  
/protein\_id="BAC29456.1"  
/db\_xref="GI:26311452"  
/translation="MERPRGTEAPSPSEMEEDDLDFGGVDSFRSYNSAGSBS  
SSYLESEAEDEDEAGELTSPHLFSSANNRSLDGSSEPAVCEMCGIVGTREAF  
FSKTRFCSVSCSSYSNSKKASILARLQKPTKKAKVLHKAASAKICAFHLAQ  
TGLADPTGQDALVGLFQWKFGLKDHYSKAAFPVSCFHVPLVDOWEDVVMKQVEV  
LNSDAVLPSRVYATVIAQAGRVLLRYEGFENDASHDFWNLGTVDVHPGLGCATN  
LTPVPRTHAKFTDKMSYLMKELVGSRTLPADFHTKMYESMKYPRFGVRLRVUDK  
TOVSTEMAVDTVIGRELILYEDGSDDFWCHMSPLIHPVGNRRRVGHGIMSD  
RCDMSHPFRKIYCDAPVLFKRVAVTEGWFEEGMKLEADILNLSIGSVATI  
CMLLDGLYLMICVDGSPSTGSDWFCVHASSHAIFPAFCQKNDIELTPPKYETOFF  
AMETYLEKTSKAPARLFNMDCFHGVGMKLEAVDLMEPLRICVATKRVVHRL  
SIHFDGNDYDQWDCESPDYIPVWCELTGYLOPPVSAEPNTPOKGDVTKKKK  
QFGKKRRIIPSAKTRPLRQSKKFLLEDNLALGVSPVDDIIAIVCVKEEHQDISL  
DRSPSPOLPLPIESIKQRNN"  
3418. 3423  
/note="putative"  
3431  
/note="putative"  
BASE COUNT 801 a 920 c 971 g 739 t  
ORIGIN  
Query Match 47.0%; Score 1200.4; DB 11; Length 3431;  
Best Local Similarity 86.4%; Pred. No. 3.4e-272;  
Matches 1351; Conservative 0; Mismatches 206; Indels 7; Gaps 2;  
QY 1 AGCAGTCCCACTCTATGACAGTGGGAGGATGTGATGAAGGATCAAGGTGAGGTGC 60  
DB 661 AACATGTACCCCTCTATGACAGTGGGAGGATGTGATGAAGGATCAAGGTGAGGTGC 720  
QY 61 TCAACAGTGTGTGTCTCCACAGCGGGTGTACTGGATCGCTCTGTCTCATCAGACAG 120  
DB 721 TCAACAGCGATGTGTCTCCACAGCGGGTGTACTGGATCGCTCTGTCTCATCAGCGAG 780  
QY 121 CAGGATTCGGGTGTCTCTCGGTGTATGAGGCTTGAATGATGACGCGGATGCTTCT 180  
DB 781 CTGGTACCGGGTGTCTCTCGGTGTATGAGGCTTGAATGATGACGCGGATGCTTCT 840  
QY 181 GGTGCAACCTGGGAAACAGTGGATGTCCACCCCATGGCTGTGGTGGCCATCAACAGCAAGA 240  
DB 841 GGTGCAACCTGGGAAACAGTGGATGTCCACCCCATGGCTGTGGTGGCCATCAACAGCAAGA 900  
QY 241 TCCTAGTGTGGGAAACAGTGGATGTCCACCCCATGGCTGTGGTGGCCATCAACAGCAAGA 300  
DB 901 TCCTAGTGTGGGAAACAGTGGATGTCCACCCCATGGCTGTGGTGGCCATCAACAGCAAGA 960  
QY 301 AACGGTGTGGGAAACAGTGGATGTCCACCCCATGGCTGTGGTGGCCATCAACAGCAAGA 360  
DB 961 AACGGTGTGGGAAACAGTGGATGTCCACCCCATGGCTGTGGTGGCCATCAACAGCAAGA 1020  
QY 361 TGAAGTACCCCTTTAGGACGGGATCGGGCTGGAAGTGTGGTGGCAAGTCCAGGTGTAC 420  
DB 1021 TGAAGTACCCCTTTAGGACGGGATCGGGCTGGAAGTGTGGTGGCAAGTCCAGGTGTAC 1080  
QY 421 GCATCTGCGATGGTGTGGTGGACACAGTAAATCGGGGTGCGCTACGCTCTCTACGAGG 480  
DB 1081 GTACCCGATGGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTACGCTCTCTACGAGG 1140  
QY 481 ATGGTGTGACGTGACGAGCTTCTGGTGGCAATGTGGAGCCCTGTATGCCACCCAGTGG 540  
DB 1141 ATGGTGTGACGTGACGAGCTTCTGGTGGCAATGTGGAGCCCTGTATGCCACCCAGTGG 1200  
QY 541 GTTGTGTACGAGCTGTGGGCGGACGGATCAAGATGTGACAGAGGCGCAAGTGCATGCC 600  
DB 1201 GTTGTGTACGAGCTGTGGGCGGACGGATCAAGATGTGACAGAGGCGCAAGTGCATGCC 1260





```

QY 1511 AACAGAGAGTGGCTTCTGCTCAGCTGCTTCTAGCTGGAGGAGCCAGCCAGCCT 1570
Db 449 AACAGAGAGTGGCTTCTGCTCAGCTGCTTCTAGCTGGAGGAGCCAGCCAGCCT 508
QY 1571 TTCTCTACCAACCAACCACTCCACCTGACTTTGGCTTGAGACTGATCTCTCTGT 1630
Db 509 TTCTCTACCAACCAACCACTCCACCTGACTTTGGCTTGAGACTGATCTCTCTGT 568
QY 1631 GTAAATTTCCCGGTGCTGTGAAGCTGGAAGCTGGAAGCTGCTGCGGTCTCTGGG 1690
Db 569 GTAAATTTCCCGGTGCTGTGAAGCTGGAAGCTGGAAGCTGCTGCGGTCTCTGGG 628
QY 1691 ACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1750
Db 629 ACCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 688
QY 1751 CCAGAACTCGTCTGGAACCACTTTTCCAGCAGAGTTCCAAAGCTGGAACCTAGC 1810
Db 689 CCAGAACTCGTCTGGAACCACTTTTCCAGCAGAGTTCCAAAGCTGGAACCTAGC 748
QY 1811 TGCTGCTCTTCTTAAAGTGGCTTCCCGCAGCCGCGCAGCCGCTGAGTTGCCAGGA 1870
Db 749 TGCTGCTCTTCTTAAAGTGGCTTCCCGCAGCCGCGCAGCCGCTGAGTTGCCAGGA 808
QY 1871 TGCGGCGCACCACTGTCACTGTGAATACAGACAGTGAACCTGTGCTGCTGGAACGAG 1930
Db 809 TGCGGCGCACCACTGTGACACTGTGAATACAGACAGTGAACCTGTGCTGCTGGAACGAG 868
QY 1931 TCATGTAAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATAGGTCCTCTGACAGTGT 1990
Db 869 CCATGTAAATTAAGTTCTAGAGAGCTCTCTGAGCAGGATAGGTCCTCTGACAGTGT 928
QY 1991 TGTGTGCTGGGCGAGCTCTGCTCCTAGCTCCCAAGGATGTTGGGACCCAGCTTGTCTCGG 2050
Db 929 TGTGTGCTGGGCGAG-CTCTGCTCCTAGCT-CCAGAGGATGTTGGGA-CCAGCTTGTCTCGG 986
QY 2051 TGTGTGCTGGGCGAGCTCTGCTCCTAGCTCCCAAGGATGTTGGGACCCAGCTTGTCTCGG 2110
Db 987 TGTGTGCTGGGCGAG-CTCTGCTCCTAGCT-CCAGAGGATGTTGGGA-CCAGCTTGTCTCGG 1043
QY 2111 CAGCTAAGAGCAGTACCAGAGTGTGATTTGGGACCTGTGTGG 2157
Db 1044 -MCHAAARAGCAGTNACMA-GATGTGATTTGGGACTGTGTGTGG 1088

RESULT 5
BX360491/c
LOCUS
DEFINITION
BX360491 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1070YF10 3-PRIME, mRNA sequence.
ACCESSION
BX360491
VERSION
BX360491.1 GI:30374437
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1065)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10409.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0D1070DC05NP1&cluster=10409.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0D1070DC05NP1.

```

```

FEATURES
source
Location/Qualifiers
1..1065
/organism="Homo sapiens"
/mo_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D1070YF10"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 251 a 270 c 300 g 218 t 26 others
ORIGIN
Query Match 36.58; Score 932.8; DB 13; Length 1065;
Best Local Similarity 94.18; Pred. No. 4e-209;
Matches 1000; Conservative 18; Mismatches 38; Indels 7; Gaps 5;
QY 1468 CAAGTCCAGAGTGCCTGTCTCCGTCCAGAACATCAAGCAGAGAAACAGAGCACTGAGCCT 1527
Db 1057 CAAGTTCAACACMRAGYGCCTTTCBCKGACATCAAGCGGAAACAGACGA--TGAGCT 1000
QY 1528 TCCTGCCCTCCAGCCTGCTTCTAGCTGGAAGCCAGCCAGCCTTCTCTACCAACCCAC 1587
Db 999 TCCTGCCCTCCAGCCTGCTTCTAGCT--GAAGCCAGCCAGCCTTCTCTACCAACCCAC 942
QY 1588 CATGCCCTCCACCTGACTTTGGCTTGAGACTGATCTCTCTGTGTAAATTTCTCCCGGTG 1647
Db 941 CATGCCCTCCACCTGACTTTGGCTTGAGACTGATCTCTCTGTGTAAATTTCTCCCGGTG 882
QY 1648 CTGTGAAGGCTGACCGGTGAGGACCTGTCTGGGTCTCTCTGGGACCCGCTGTGTCTTCT 1707
Db 881 CTGTGAAGGCTGACCGGTGAGGACCTGTCTGGG--TCTCTGGACCCGCTGTGTCTTCT 824
QY 1708 GCCTCTCCCTGTGAAAGTCTATATGACGGGCGGCTGAGGCCCCAGAACTCGTCTGTG 1767
Db 823 GCCTCTCCCTGTGAAAGTCTATATGACGGGCGGCTGAGGCCCCAGAACTCGTCTGTG 764
QY 1768 AACCACTTTTCCAGCCAGAGTTCCCAAAGCTGGAAAGCTAGCTGCTCTCTTCTTAA 1827
Db 763 AACCACTTTTCCAGCCAGAGTTCCCAAAGCTGGAAAGCTAGCTGCTCTCTTCTTAA 704
QY 1828 GATGGCTCTCCCGGACCGCCAGCCGCTCAGTTGCCAGGATGGGGCCACCAGCTGTCA 1887
Db 703 GATGGCTCTCCCGGACCGCCAGCCGCTCAGTTGCCAGGATGGGGCCACCAGCTGTCA 644
QY 1888 CACTGTGGAATACAGACAGTGAACCTCTGCTGCTGCTGAACAGAGTCAATGTAATTAAGTTC 1947
Db 643 CACTGTGGAATACAGACAGTGAACCTCTGCTGCTGCTGAACAGAGCCATGTAATTAAGTTC 584
QY 1948 TAGAGCAGCTCTCTGAGCAGGATTAAGGTCCTCTGACAGTGTGTGTGGTGGGGCAGC 2007
Db 583 TAGAGCAGCTCTCTGAGCAGGATTAAGGTCCTCTGACAGTGTGTGTGGTGGGGCAGC 524
QY 2008 CTCCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCCTCATGTTGTGGTCTCTGCT 2067
Db 523 CTCCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCCTCATGTTGTGGTCTCTGCT 464
QY 2068 CTCCTGCTCAAAAATTCACCAAGCAGATGTTGGGAGCCAGCTTGTCTCGGAGCTAAGAAAGCAGTGA 2127
Db 463 CTCCTGCTCAAAAATTCACCAAGCAGATGTTGGGAGCCAGCTTGTCTCGGAGCTAAGAAAGCAGTGA 404
QY 2128 CCAGGATGGGATTTTGGGACCTGTGTGGGCTTGTAGCTGCTTCTGTGTGTGTGAG 2187
Db 403 CCAGGATGGGATTTTGGGACCTGTGTGGGCTTGTAGCTGCTTCTGTGTGTGTGAG 344
QY 2188 GACTGACTCCCAATTTCTTAAAGGAAATCCCGGGGAGGACATTGGGAGGAGATGGCC 2247
Db 343 GACTGACTCCCAATTTCTTAAAGGAAATCCCGGGGAGGACATTGGGAGGAGATGGCC 284
QY 2248 TGAGTGTGAGCTTTGGCTCTGCTGCTGCTGAGGCCCGCTTAAATAATTAATTCATCA 2307

```

```

Db      283 TGAGTGTGCACTTTGGCTCTGCTACCTGCTCTGTAAGAGCCCGCTAAAAATAATCAATCCA 224
QY      2308 AGATTCCTTTTGTAGTAAAGGGTCCAGATTTCTGACTGGAGCCTCTAGAGAGCTGGGCTGTG 2367
Db      223 AGATTCCTTTTGTAGTAAAGGGTCCAGATTTCTGACTGGAGCCTCTAGAGAGCTGGGCTGTG 164
QY      2368 ATGTTCTTTTGGCTTTTGTCTTACTAAATGAAGAAACCATGCTGGAGGGCCGTGA 2427
Db      163 ATGTTCTTTTGGCTTTTGTCTTACTAAATGAAGAAACCATGCTGGAGGGCCGTGA 104
QY      2428 ACACAGAACCTTCAAGACAAAGATGACAGAGCTGGAGGACACATCTAGCTGCCATTGCAA 2487
Db      103 ACACAGAACCTTCAAGACAAAGATGACAGAGCTGGAGGACACATCTAGCTGCCATTGCAA 44
QY      2488 CC-TGACTGGGCTCCCGAGCTCTGTGTGTGAGAAATTAACC 2529
Db      43 CCNNCACTGGGCTCCGSGGACTGTGTGTCTNNNDKTRRACS 1

RESULT 6
AL560108
LOCUS
DEFINITION
AL560108 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
ACCESSION
VERSION
AL560108.2 GI:31284239
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 15, 2001 this sequence version replaced gi:12906251.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10409.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DG002BC07QP1&cluster=10409.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DG002BC07QP1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG002YE14"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."
BASE COUNT 259 a 294 c 313 g 288 t 47 others
ORIGIN
Query Match 34.4%; Score 878; DB 9; Length 1201;
Best Local Similarity 99.8%; Pred. No. 3,4e-196;
Matches 900; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY      1643 CGGTGCTGTGAGAGCTGGAGCTGGAGAGCTGCTGGGGTCTCTGGAGCCCGCTGTG 1702
Db      68 CGGTGCTGTGAGAGCTGGAGCTGGAGAGCTGCTGGGGTCTCTGGAGCCCGCTGTG 127

```

```

QY      1703 CTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGGCGCCTGAGGCCCCAGAACTCGT 1762
Db      128 CTTCTGCCCTCCCTGTGGAAGGTCTATATGACGGGCGCCTGAGGCCCCAGAACTCGT 187
QY      1763 CTGTGAACCACTTTTTCAGCCAGAGTTCCCAAGAGTGGAAACGCTAGCTGCCTGCTCTTC 1822
Db      188 CTGTGAACCACTTTTTCAGCCAGAGTTCCCAAGAGTGGAAACGCTAGCTGCCTGCTCTTC 247
QY      1823 CTTTAAAGATGCTCCCTCCCGACCGGCAACCGCCCTCAGTTGCCAGGATGGGGCCACCAC 1882
Db      248 CTTTAAAGATGCTCCCTCCCGACCGGCAACCGCCCTCAGTTGCCAGGATGGGGCCACCAC 307
QY      1883 TGTCAACTGTGGAATACAGACAGTGAACCTGTGTCTGCTGAAACGAGTCATGTAATTA 1942
Db      308 TGTCAACTGTGGAATACAGACAGTGAACCTGTGTCTGCTGAAACGAGTCATGTAATTA 367
QY      1943 AGTTCTAGAGCAGCTCTCTGAGCAGGATAGGTCCCTGACAGTGAAGTTGTGTGGGG 2002
Db      368 AGTTCTAGAGCAGCTCTCTGAGCAGGATAGGTCCCTGACAGTGAAGTTGTGTGGGG 427
QY      2003 GCAGCCTCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCCCTCATGTGTTGCTCT 2062
Db      428 GCAGCCTCTGCTCAAAAATTCACCAAGCAGATGCTCTCAGCCCTCATGTGTTGCTCT 487
QY      2063 CTGCTCCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTGGCAGCTAAGAAC 2122
Db      488 CTGCTCCTCTAGCTCCCGAGGATGTTGGGACCCAGCTTGTCTGGCAGCTAAGAAC 547
QY      2123 AGTGACACAGATGTGATTTTGGCAGCTGTGTGGGCTTGGCTGTGTTCTGTGTTT 2182
Db      548 AGTGACACAGATGTGATTTTGGCAGCTGTGTGGGCTTGGCTGTGTTCTGTGTTT 607
QY      2183 GTGAGGACTGACTCCCATTTCTTAAAGGAAATGCCCGGGGAGGACATTTGGGAGAGA 2242
Db      608 GTGAGGACTGACTCCCATTTCTTAAAGGAAATGCCCGGGGAGGACATTTGGGAGAGA 567
QY      2243 TGGCCTGAGTGTGCACTTTGGCTCTGCTACCTGCTCTTGAAGCCCGCTGAAATAATTC 2302
Db      568 TGGCCTGAGTGTGCACTTTGGCTCTGCTACCTGCTCTTGAAGCCCGCTGAAATAATTC 727
QY      2303 ATCCAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCTCTAGAGAGCTGG 2362
Db      728 ATCCAAGATTCCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCTCTAGAGAGCTGG 787
QY      2363 CTTGTATGTTCTTTTGGCTTTTGGCTTCTTAAATGAAGAAACCATGCTGGAGGGGC 2422
Db      788 CTTGTATGTTCTTTTGGCTTTTGGCTTCTTAAATGAAGAAACCATGCTGGAGGGGC 847
QY      2423 CGTGAACACAGAACCTTCAAGACAGGATGACAGAGCTGGAGGACACATCTAGCTGCCAT 2482
Db      848 CGTGAACACAGAACCTTCAAGACAGGATGACAGAGCTGGAGGACACATCTAGCTGCCAT 907
QY      2483 TGCACCTCCTAGTGGGCTCCCGAGCTCTGTGTGTGAGAAATTAACCCCTGCTGTGTTG 2542
Db      908 TGCACCTCCTAGTGGGCT-CCGAGACTCTGTGTGTGAGAAATTAAC-CCCTGCTGTGTTG 965
QY      2543 AG 2544
Db      966 AG 967

RESULT 7
BM909451
LOCUS
DEFINITION
AGENCOURT_6640956 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5434030
5', mRNA sequence.
ACCESSION
BM909451
VERSION
BM909451.1 GI:19359830
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```



laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC Library."

BASE COUNT 194 a 286 c 249 g 171 t

ORIGIN

Query Match 31.1%; Score 794.6; DB 13; Length 900;  
Best Local Similarity 98.2%; Pred. No. 1.6e-176;  
Matches 814; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 1189 AGCTACCGGCTACAGCTCCAGCTCCCTGTGTGCGCAGAACCGGCGCCACCGCTGAAGG 1248  
DB 1 AGCTACCGGCTACAGCTCCAGCTCCCTGTGTGCGCAGAACCGGCGCCACCGCTGAAGG 60

QY 1249 CCAAGAGGCGCCAAAGAGAAAGAAAGAAAGTTGGGAGAAAGAAAGAAAGAAATCCCGC 1308  
DB 61 CCAAGAGGCGCCAAAGAGAAAGAAAGAAAGTTGGGAGAAAGAAAGAAAGAAATCCCGC 120

QY 1309 CCACTAAGAGCGGACCCCTCAGACAGGGGTCCAGAGAGCCCTGTGTGAGGAGCAGCCCTC 1368  
DB 121 CCACTAAGAGCGGACCCCTCAGACAGGGGTCCAGAGAGCCCTGTGTGAGGAGCAGCCCTC 180

QY 1369 AGGCTGCCAGGAGATCTGTGCGAGCTGTTCCTGCGGAGATCAATTCGTGCGTGTGA 1428  
DB 181 AGGCTGCCAGGAGATCTGTGCGAGCTGTTCCTGCGGAGATCAATTCGTGCGTGTGA 240

QY 1429 AGAAGAGCATCTAGACGTGGCTCCCGCAGACAGGCTTCAAGTCACAGAGTCGCTGCT 1488  
DB 241 AGAAGAGCATCTAGACGTGGCTCCCGCAGACAGGCTTCAAGTCACAGAGTCGCTGCT 300

QY 1489 CGTCCAGAAATCAAGCAGGAAAGCAGACGATGAGCTTCTGCTCCAGCTGGCTTC 1548  
DB 301 CGTCCAGAAATCAAGCAGGAAAGCAGACGATGAGCTTCTGCTCCAGCTGGCTTC 360

QY 1549 TAGCTGGAAGCCAGCCAGCGTTCTCTACACACACACACACCTGCTCCAGCTGATTTGG 1608  
DB 361 TAGCTGGAAGCCAGCCAGCGTTCTCTACACACACACACACCTGCTCCAGCTGATTTGG 420

QY 1609 CTTGAGAGCTGATCTCTGTGTAATTTCTGCCGGTGTGTGAAAGCTGAGCGGTGA 1668  
DB 421 CTTGAGAGCTGATCTCTGTGTAATTTCTGCCGGTGTGTGAAAGCTGAGCGGTGA 480

QY 1669 GAGCTGTGGGGTCTCTGGGAGCGGCTGTGTTCTGCTCCCTCTGTTGAAAGGTC 1728  
DB 481 GAGCTGTGGGGTCTCTGGGAGCGGCTGTGTTCTGCTCCCTCTGTTGAAAGGTC 540

QY 1729 TATATGACGGGCGGCTGAGGCGCCAGAACTCGTGTGAAACCACTTTTCCAGCCAGAG 1788  
DB 541 TATATGACGGGCGGCTGAGGCGCCAGAACTCGTGTGAAACCACTTTTCCAGCCAGAG 600

QY 1789 TTCCCAAGAGCTGGAAGCTAGTGTGCTGCTCTTCTTAAGATGGCTTCCCGGACCGC 1848  
DB 601 TTCCCAAGAGCTGGAAGCTAGTGTGCTGCTCTTCTTAAGATGGCTTCCCGGACCGC 660

QY 1849 CACGGCCCTCAGTTGCCAGGATGGGGCCACCACTGTCCACTGTGGAATACAGACAGT 1908  
DB 661 CACGGCCCTCAGTTGCCAGGATGGGGCCACCACTGTCCACTGTGGAATACAGACAGT 720

QY 1909 GAACCTGTGCTGTGAACGATGATGTAATTAAGTTCTAGAGCAGCTCTCTGAGCAGG 1968  
DB 721 GAACCTGTGCTGTGAACGATGATGTAATTAAGTTCTAGAGCAGCTCTCTGAGCAGG 780

QY 1969 ATAA-GGTCCCTGACAGTGTGTGGTGGGGGAGCTCTGCGCTC 2016  
DB 781 ATAA-GGTCCCTGACAGTGTGTGGTGGGGGAGCTCTGCGCTC 829

RESULT 9  
BQ927459  
LOCUS  
DEFINITION BQ927459 927 bp mRNA linear EST 20-AUG-2002  
AGENCOURT 8822242 NIH\_MGC\_40 Homo sapiens cDNA clone IMAGE:6381422  
5', mRNA sequence.

BQ927459  
BQ927459.1 GI:22342490  
EST.  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NIH-MGC http://mgc.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs@mail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Rubin Laboratory  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone distribution: Agencourt Bioscience Corporation  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM2571 row: j column: 15  
High quality sequence stop: 620.

FEATURES  
source  
1..927  
Location/Qualifiers  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6381422"  
/tissue\_type="carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 40"  
/note="Organ: prostate; Vector: pOTB7; Site 1: XhoI;  
Site 2: EcoRI; CDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGACGAG(G). Library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH MGC Library." 2 others

BASE COUNT 194 a 254 c 251 g 226 t

ORIGIN

Query Match 30.6%; Score 783; DB 13; Length 927;  
Best Local Similarity 96.9%; Pred. No. 8.6e-174;  
Matches 809; Conservative 0; Mismatches 22; Indels 4; Gaps 1;

QY 1625 CTCTGTGTAATTTCTGCCGGTGTGTGAAAGCTGAGCGTGGAGGACCTGCTGGGGTCT 1684  
DB 1 CTCTGTGTAATTTCTGCCGGTGTGTGAAAGCTGAGCGTGGAGGACCTGCTGGGGTCT 60

QY 1685 CTTGGACCCCGCTGTGTTCTTCTGCCCTCCCTGTGAAAGGTCATATGACGGGCCGCC 1744  
DB 61 CTTGGACCCCGCTGTGTTCTTCTGCCCTCCCTGTGAAAGGTCATATGACGGGCCGCC 120

QY 1745 TGAGGCCCCAGAACCTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTGGAAC 1804  
DB 121 TGAGGCCCCAGAACCTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTGGAAC 180

QY 1805 GCTAGCTGCTGCTCTTCTTAAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTTCCTCTGCA 1864  
DB 181 GCTAGCTGCTGCTCTTCTTAAAGTTCCTAGACAGCTCTCTGAGCAGGCTCTGAGTTC 240

QY 1865 CAGGATGGGGCCACCACTGTGCACTGTGGAATACAGACAGTGAATCTGTCTGCTG 1924  
DB 241 CAGGATGGGGCCACCACTGTGCACTGTGGAATACAGACAGTGAATCTGTCTGCTG 300

QY 1925 AACGAGTCATGTAAATTAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTTCCTCTGCA 1984  
DB 301 AACGAGTCATGTAAATTAAGTTCCTAGACAGCTCTCTGAGCAGGATAGTTCCTCTGCA 360

QY 1985 GTGAGTGTGTGGTGGGGCAGCTCTGCTCAAAATTTACCAAGCAGAGATGCTCTCA 2044  
DB 361 GTGAGTGTGTGGTGGGGCAGCTCTGCTCAAAATTTACCAAGCAGAGATGCTCTCA 420





Db	481	TCGAGACTCTTTAACTGATGGATTCGCCAAACCATGCGTTCGAGGTGGCATGAACTGGAG	540		
QY	1035	GCCTGTGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGAACAGAGTGGTGCAT	1094		
Db	541	GCCTGTGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGAACAGAGTGGTGCAT	600		
QY	1095	CGGCTCTCAGCATCCACTTTTGACGGCTGGGACAGCAGTACGACCACTGGGTGGACTGC	1154		
Db	601	CGGCTCTCAGCATCCACTTTTGACGGCTGGGACAGCAGTACGACCACTGGGTGGACTGC	660		
QY	1155	GAGTCCCGACACATCTACCCCGTGGGTGGTGTGAGCTCACC-GGCTACCAAGTCC-AGC	1212		
Db	661	GAGTCCCGACACATCTACCCCGTGGGTGGTGTGAGCTCACC-GGCTACCAAGTCC-AGC	720		
QY	1213	CTCTGTGGCGCGAG-AACCGGGCCACACCGCTGAAGG--CCAAAGAGGCCACAAAGAAGA	1269		
Db	721	CTCTGTGGCGCGAGAAACCGGGCCACACCGCTGAAGGGCCAAAGAGGGCCCCNNAGAAGA	780		
QY	1270	AAAGAAGACAGTTTGGG--AAGAAAGGAAAGAGATCCGCCCACTAAGACGCGACCCC	1326		
Db	781	AAAGAAGACCGTTTGGGAAAGAAAGGAAAAAATCCGCCCACTAGGAAGCGGACCC	840		
RESULT 12	BI092956	851 bp	mrna	linear	EST 20-JUN-2001
LOCUS	602857991.F1 NIH_MGC_10	Homo sapiens	cDNA clone IMAGE:499294	5'	
DEFINITION	mrna sequence.				
ACCESSION	BI092956				
VERSION	BI092956				
KEYWORDS	EST.				
SOURCE	BI092956.1 GI:14511286				
ORGANISM	Homo sapiens (human)				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
JOURNAL	1 (bases 1 to 851)				
COMMENT	NIH-MGC http://mgc.nci.nih.gov/ National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM1029 row: e column: 23 High quality sequence stop: 803. Location/Qualifiers 1..851 /organism="Homo sapiens" /mol_type="mrna" /db_xref="taxon:9606" /clone="IMAGE:499294" /cell_line="MGC36" /lab_host="DH10B" /clone_lib="NIH_MGC 10" /note="Organ: cervix; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo df. Average insert size 1.5 kb. Library prepared by Life Technologies."				
BASE COUNT	182 a	234 C	231 G	204 T	
ORIGIN					
Query Match	29.0%	Score 739.8;	DB 12;	Length 851;	
Best Local Similarity	97.5%	Pred. No. 1.3e-163;			
Matches 836;	Conservative	0;	Mismatches 12;	Indels 9;	Gaps 8;



Db	1	AGGACCTGCTGGGGTCTCTGGGACCGCGCTGTGCTTCTGCCCTCCCTGTGGAAGGT	60
Qy	1728	CTATATGACGGGCGCGCTGAGGCGCCAGAACTCGTCTGTGGAACACACTTTTCAGCCAGA	1787
Db	61	CTATATGACGGGCGCGCTGAGGCGCCAGAACTCGTCTGTGGAACACACTTTTCAGCGAGA	120
Qy	1788	GTTCCCAAAGCTGGAACGCTAGTGCCTGTCTTCTTAAAGTGGCTCCCCCGACCCG	1847
Db	121	GTTCCCAAAGCTGGAACGCTAGTGCCTGTCTTCTTAAAGTGGCTCCCCCGACCCG	180
Qy	1848	CCACGGGCGCTCAGTTGCTACAGGGATGGGGCCACCACTGTCTACATGTGGATACAAGACAG	1907
Db	181	CCACGGGCGCTCAGTTGCTACAGGGATGGGGCCACCACTGTCTACATGTGGATACAAGACAG	239
Qy	1908	TGAACCTGTGCTGCCTGAAACGAGTCATGTAAATTAAGTTCTAGACAGAGTCTCTGAGCAG	1967
Db	240	TGAACCTGTGCTGCCTGAAACGAGTCATGTAAATTAAGTTCTAGACAGAGTCTCTGAGCAG	299
Qy	1968	GATAAGTCCCTGACAGTGAAGTTGTGTGGTGGGGCAGCCTCTGCTCAAAATTCACC	2027
Db	300	GATAAGTCCCTGACAGTGAAGTTGTGTGGTGGGGCAGCCTCTGCTCAAAATTCACC	359
Qy	2028	AAGCAGAATGCCCTCTCAGCCCTCATGTGTGGTCTCTGCTCTGCTCTCTTCCCTAGCCAGGAT	2087
Db	360	AAGCAGAATGCCCTCTCAGCCCTCATGTGTGGTCTCTGCTCTCTTCCCTAGCCAGGAT	419
Qy	2088	GTTGGGGAACCGACTGTCTTGGCAGCTAAGAACAGTGAACAGGATGTGGATTTTGGCG	2147
Db	420	GTTGGGGAACCGACTGTCTTGGCAGCTAAGAACAGTGAACAGGATGTGGATTTTGGCG	479
Qy	2148	ACCTGTGTGGTGGCCCTGAGCTGCTTTCCTGTGTGTGTGAGGAGTGAATCCCATTTCTTAA	2207
Db	480	ACCTGTGTGGTGGCCCTGAGCTGCTTTCCTGTGTGTGTGAGGAGTGAATCCCATTTCTTAA	539
Qy	2208	AGGAATGCCCCGGGGAGGACATTGGGAGGAGATGGCCTGAGTGTGCACATTTGGCTCT	2267
Db	540	AGGAATG-CCCCGGGGAGGACATTGGGAGGAGATGGCCTGAGTGTGCACATTTGGCTCT	598
Qy	2268	GCTACTGCTCTGTAAGCCCCGCTTAA-NATAATTCATCCAGATTCCTTTGTAGTTAAA	2326
Db	599	GCTACTGCTCTGTAAGCCCCGCTTAA-NATAATTCATCCAGATTCCTTTGTAGTTAAA	658
Qy	2327	GGGT-CCAGTTCTGACTGAGGCTCTAGACAGCTGGGCTTGTATGTTCTTTTGGCTTTT	2385
Db	659	GGGTGCCAGTTCTGATTGGAGCCTCTAGAGACTGGGCTTGTATGTTCTTTTGGCTTTAG	717
Qy	2386	GTTCTTACTTAATGAAGAACCATGCTTGGAGGGGCGGTGAACACAGAACCTCAAGAC	2445
Db	718	GTTCTTACTTAATGAAGAACCATGCTTGGAGGGGCGGTGAACACAGAACCTCAAGAC	777
Qy	2446	AAGGATGA-CAGAGCTGGAGGACACATCTAGCTGCGCATGTCAACTCTACTGGGCTCCCCA	2504
Db	778	AAGGATGACCAGAGCTGCAGGACACAT-TAAGTGCAT--GCAGCTCACTGGGCTCCCCA	834
Qy	2505	GACTCTGTGTGAGAA	2521
Db	835	GACTCTGTGTGAGAA	851

RESULT 13  
BG396658  
LOCUS  
DEFINITION  
602459726F1 NIH MGC\_16 Homo sapiens cDNA clone IMAGE:4581846 5',  
mRNA sequence.  
1057 bp mRNA linear EST 12-MAR-2001  
BG396658  
ACCESSION  
BG396658  
VERSION  
BG396658.1 GI:13290106  
KEYWORDS  
EST.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1057)  
REFERENCE  
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">image.llnl.gov</a> Plate: LLC1M1303 row: d column: 07 High quality sequence stop: 763.

**FEATURES**  
**SOURCE**

```

1. .1057
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4581846"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_16"
Notes:"Organ: eye; Vector: pOTF7; S
ECORI; cDNA made by oligo-dT primin
cloned into EcoRI/XhoI sites using
adapter: GGCACGAG(G). Library cons
in the laboratory of Gerald M. Rubin
California, Berkeley) using ZAP-CDN
(Stratagene) and Superscript II RT
Note: this is a NIH_MGC Library."
254 a 236 c 303 g 203t 1
BASE COUNT

```

BASE COUNT 254 a 296 c 303 g 203 t 1 others

ORIGIN

Note: this is a NIH\_MGC Library."

Query Match	28.9%	Score 737.6	DB 10	Length 1057	
Best Local Similarity	93.4%	Pred. No. 4.5e-163			
Matches 825	Conservative	0	Mismatches 5	Indels 8	Gaps 5
QY	1247	GGCCAAAGAGCGCCACAAAGAGAAAAGAAACAGATTGGGRAGAAAAGAAAAGATCCC	1306		
Db	1	GGCCAAAGAGCGCCACAAAGAGAAAAGAGACAG-TTGGGAGAAAAGAAAAGAAATCCC	59		
QY	1307	GCCCACTAAGACGCGACCCCTCAGACAGGGGTCCAAGAGAGCCCTGCTGGAGGACGACCC	1366		
Db	60	GCCCACTAAGACGCGACCCCTCACAGGGGTCCAAGAGAGCCCTGCTGGAGGACGACCC	115		
QY	1367	TCAGGTCGCCAGAGAATCTGTCGAGAGCTTTTCTCTGGCAGATCAATGCTGTGCGTGT	1426		
Db	120	TCAGGTCGCCAGAGAATCTGTCGAGAGCTTTTCTCTGGCAGATCAATGCTGTGCGTGT	179		
QY	1427	GAAGGAAGACATCTAGACGTGSCCTCCGCCACAGAGCTTTCAAGTCCAGAGCTGCCTGT	1486		
Db	180	GAAGGAAGACATCTAGAGCTGGCTTCGCCGACAGAGCTTTCAAGTCCAGAGCTGCCTGT	239		
QY	1487	CTCGTTCGAGAACATCAAGCAGGAGAAACAGACGACTGAGCCTTCTCTGCTCCAGCCTCGCT	1546		
Db	240	CTCGTTCGAGAACATCAAGCAGGAGAAACAGACGACTGAGCCTTCTCTGCTCCAGCCTCGCT	299		
QY	1547	TCTAGCTGGAGCCAGCCAGCGTTTCTCTACACACACACATGCTTCACCTGACCTTT	1606		
Db	300	TCTAGCTGGAGCCAGCCAGCGTTTCTCTACACACACACATGCTTCACCTGACCTTT	359		
QY	1607	GGCTTGGAGACTGATCCTCTCTGTGTAATTTCTGCCGGTCTCTGAGAGCTGACGCGTG	1666		
Db	360	GGCTTGGAGACTGATCCTCTCTGTGTAATTTCTGCCGGTCTCTGAGAGCTGACGCGTG	419		
QY	1667	GAGACCTGCTGGGTCCTCTGGGACCGCCTGTTGCTTCTGCCCCTCCCTGTGGAAAGG	1726		
Db	420	GAGAGCTGCTGGGTCCTCTGGGACCGCCTGTTGCTTCTGCCCCTCCCTGTGGAAAGG	479		
QY	1727	TCTATATACGGCGCGCTGAGGCCCAAGAACTGCTGTGTGAACCACTTTTCAGCCAG	1786		
Db	480	TCTATATACGGCGCGCTGAGGCCCAAGAACTGCTGTGTGAACCACTTTTCAGCCAG	539		
QY	1787	AGTTCACCAAGCTGGAAAGCTAGTGTGCTCTTCTCTTAAGATGGCTCTCCCCCGACCC	1846		



ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 855)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: csapbs-remail.nih.gov  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LHM1618 row: d column: 16  
 High quality sequence stop: 2  
 High quality sequence stop: 781.  
 FEATURES  
 source  
 1..855  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clones="IMAGE:5245023"  
 /lab\_host="DH10B"  
 /clone\_lib="NIH\_MGC\_121"  
 /note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI;  
 Site\_2: EcoRV (destroyed); RNA source anonymous pool of 3  
 fetal brains, female age 20 weeks, female age 24 weeks,  
 and male age 26 weeks. Library is oligo-dr primed and  
 directionally cloned (EcoRV site is destroyed upon  
 cloning). Average insert size 1.7 kb, insert size range  
 0.7-3.5 kb. Library is normalized and enriched for  
 full-length clones and was constructed by C. Gruber  
 (Invitrogen). Research Genetics tracking code 017. Note:  
 this is a NIH\_MGC Library."  
 BASE COUNT 186 a 258 c 245 g 166 t  
 ORIGIN  
 Query Match 28.7%; Score 733.4; DB 12; Length 855;  
 Best Local Similarity 97.0%; Pred. No. 4.3e-162;  
 Matches 83; Conservative 0; Mismatches 16; Indels 10; Gaps 8;  
 QY 982 TCCTTAACTGATGCCCCAAACCATGGCTTCAAGTGGGCGCATGAAGCTGGAGCGCGTGG 1041  
 Db 5 TCCTTAACTGATGCCCCAAACCATGGCTTCAAGTGGGCGCATGAAGCTGGAGCGCGTGG 64  
 QY 1042 ACCTGATGGAGCCCGGCTCATCTGTGTGCGCCACGGGTGAACGAGTGGTGCATCGGCTCC 1101  
 Db 65 ACCTGATGGAGCCCGGCTCATCTGTGTGCGCCACGGGTGAACGAGTGGTGCATCGGCTCC 124  
 QY 1102 TCAGCATCCACTTTGACGGGTGGGACAGCGAGTACGACCAAGTGGGTGGACTCGGAGTCCC 1161  
 Db 125 TCAGCATCCACTTTGACGGGTGGGACAGCGAGTACGACCAAGTGGGTGGACTCGGAGTCCC 184  
 QY 1162 CAGACATCTACCCCGTGGGT 1221  
 Db 185 CAGACATCTACCCCGTGGGT 244  
 QY 1222 CCGCAGAACCGGCGCACACCGCTCAAGGCCCAAGAGGCCCAAGAGGAAAGAAAGAACAGT 1281  
 Db 245 CCGCAGAACCGGCGCACACCGCTCAAGGCCCAAGAGGCCCAAGAGGAAAGAAAGAACAG- 303  
 QY 1282 TTGGGAAGAAAGAAAGAAAGATCCCGCCCACTAAGACGCGACCCCTCAGACAGGGGTCCA 1341  
 Db 304 TTGGGAAGAAAGAAAGAAAGATCCCGCCCACTAAGACGCGACCCCTCAGACAGGGGTCCA 363  
 QY 1342 AGAAGCCCTGTGTGGAGGACGACCTCAGGGTCCGAGAGAGTCTCGTCGGAGCCCTGTC 1401  
 Db 364 AGAAGCCCTGTGTGGAGGACGACCTCAGGGTCCGAGAGAGTCTCGTCGGAGCCCTGTC 422  
 QY 1402 CTGGCGAGATCAATTGCTGTGCGGTGTGAAGGAAGAGAGATCTAGACGTGGCCCTCGCCCGACA 1461

Db 423 CTGGCGAGATCAATTGCTGTGCGGTGTGAAGGAAGAGATCTAGACGTGGCCCTCGCCCGACA 482  
 QY 1462 AGGCTTCAAGTCCAGAGCTGCGTGTCTCGGTGAGAAATCATCAGCAGGAACAGACGACT 1521  
 Db 483 AGGCTTCAAGTCCAGAGCTGCGTGTCTCGGTGAGAAATCATCAGCAGGAACAGACGACT 542  
 QY 1522 GAGCCTTCTGCTCCAGCTGCGTGTCTAGTGGAAAGCCAGCCAGCCGTTTCTCTACCA- 1581  
 Db 543 GAGCCTTCTGCTCCAGCTGCGTGTCTAGTGGAAAGCCAGCCAGCCGTTTCTCTACCA- 601  
 QY 1582 CACCAACCATGCTCCAGCTGCGTGTCTAGTGGAAAGCCAGCCAGCCGTTTCTCTACCA- 1641  
 Db 602 CACCAACCATGCTCCAGCTGCGTGTCTAGTGGAAAGCCAGCCAGCCGTTTCTCTACCA- 661  
 QY 1642 CCGGTGCTGTG-AAGGCTGGAACGGTGGAGGACCTGCTGGGGTCTCTCGGACCCGCGCTGT 1700  
 Db 662 CCGGTGCTGTGAAAGGCTGGAGGACCTGCTGGGGTCTCTCGGACCCGCGCTGT 721  
 QY 1701 TGTCTTCTGCTCCCGCTCTGTGAAAGGCTCTATATGACGGGCGGCTGAGGCGCCAGAACTC 1760  
 Db 722 TGTCTTCTGCTCCCGCTCTGTGAAAGGCTCTATATGACGGGCGG-CTGAGGCGCCAGAACTC 780  
 QY 1761 GTC-TGTGAACCAACCTTTTCCAGCCAGAGTTCCCAAGCTGGAAGGCTAGCTGCTGCTC 1819  
 Db 781 GTCGTGTGAACCACTTTTCCAGCCAGAGTCCC---AAGCTGGACGCTAGCTGCTGCTC 837  
 QY 1820 TTCCTTAAAGTGGGCTCCC 1838  
 Db 838 -TCCTTAAAGTGGGCTCCC 855

Search completed: February 4, 2004, 17:47:27  
 Job time : 5231 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 13:26:23 ; Search time 673 Seconds  
(without alignments)  
10248.238 Million cell updates/sec

Title: US-10-031-915-90

Perfect score: 2555

Sequence: 1 agcagctccactatgac.....ttgotttgagaaaaa 2555

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_19Jun03.\*  
1: /SID1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*  
25: /SID1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2555	100.0	2555	22	AAFS9625 Human cell cycle a
2	2550.8	99.8	2781	22	AA158732 Human polynucleoti
3	2544	99.6	3530	25	ACCS0583 Human secreted pro
4	2544	99.6	3530	25	ABZ71320 Secreted protein-e
5	2514.4	98.4	2663	24	ABL99857 Human polynucleoti
6	2374.2	92.9	2380	22	AAI60518 Human polynucleoti
7	1434.6	56.1	3071	24	ABA3718 Human transmembran
8	413.8	16.2	3590	21	AAZ94124 Haematopoietic ste

9	405.2	15.9	2564	25	ABX08841	Angiogenesis-assoc
10	324.4	12.7	476	24	ABL77723	Human ovarian canc
11	281.4	11.0	418	24	ABL68929	Kidney cancer rela
12	271.8	10.6	289	20	AAH87128	Human single nucle
13	266.4	10.4	3486	23	ABL27733	Drosophila melanog
14	266.4	10.4	4379	23	ABL19885	Drosophila melanog
15	252.8	9.9	290	16	AAAT23623	Human gene signatu
16	250.6	9.8	251	19	AAI12386	Human biallelic po
17	235.4	9.2	558	22	ABA63686	Human foetal liver
18	235.4	9.2	558	22	ABA30876	Probe #9342 for ge
19	235.4	9.2	558	22	AAK12202	Human brain expres
20	235.4	9.2	558	22	AAK37924	Human bone marrow
21	235.4	9.2	558	22	AAI18685	Probe #8618 for ge
22	235.4	9.2	558	22	AAI43798	Probe #12484 used
23	235.4	9.2	558	23	ABS37546	Human liver single
24	235.4	9.2	558	24	ABS11920	Human genome-deriv
25	232	9.1	232	22	ABA75996	Human foetal liver
26	232	9.1	232	22	ABA40556	Probe #19022 for g
27	232	9.1	232	22	AAK24673	Human brain expres
28	232	9.1	232	22	AAK50672	Human bone marrow
29	232	9.1	232	22	AAI27682	Probe #17815 for g
30	232	9.1	232	22	AAI56650	Probe #25336 used
31	232	9.1	232	23	ABS50269	Human liver single
32	232	9.1	232	24	ABS24144	Human genome-deriv
33	167	6.5	1548	23	AAJ78585	DNA encoding novel
34	158.8	6.2	8615	23	ABL27732	Drosophila melanog
35	158.8	6.2	8670	23	ABL19884	Drosophila melanog
36	137.8	5.4	2922	20	AAAT27923	Rat l(3)mbt protei
37	135.8	5.3	165	22	AAAT25416	Human ovarian PCR-
38	133.6	5.2	3394	20	AAAT27920	Human l(3)mbt prot
39	133.6	5.2	3512	20	AAAT27921	Human l(3)mbt prot
40	133.4	5.2	4505	25	ABX34736	Human mddt cDNA SE
41	133.4	5.2	4519	25	ABX34823	Human mddt cDNA SE
42	129.4	5.1	2714	20	AAAT27922	Mouse l(3)mbt prot
43	121	4.7	555	22	ABA63649	Human foetal liver
44	121	4.7	555	22	ABA30842	Probe #9308 for ge
45	121	4.7	555	22	AAK12168	Human brain expres

## ALIGNMENTS

RESULT 1  
AAFS9625

ID AAFS9625 standard; cDNA; 2555 BP.

XX AAFS9625;

AC AAFS9625;

XX 24-APR-2001 (first entry)

DT Human cell cycle and proliferation protein CCYPR-36 cDNA, SEQ ID NO:90.

XX Cell cycle and proliferation protein; CCYPR; human; agonist;

DE antagomist; gene therapy; detection; gene therapy;

XX transgenic animal disease model; immune disorder;

DE developmental disorder; cell signalling disorder;

XX cell proliferative disorder; cancer; tumour; anaemia; epilepsy;

DE arteriosclerosis; asthma; allergy; diabetes mellitus;

XX menstrual cycle disorder; bacterial infection; ss.

OS Homo sapiens.

XX WO200107471-A2.

PN 01-FEB-2001.

PD 21-JUL-2000; 2000WO-US19948.

XX 21-JUL-1999; 99US-0145075.

PR 08-SEP-1999; 99US-0153129.

PR 10-NOV-1999; 99US-0164647.

XX (INCY-) INCYTE GENOMICS INC.

XX Hillman JL, Lal P, Tang YT, Yue H, Au-Young J, Bandman O;  
PI Azimzai Y, Yang J, Lu DAM, Baughn MR, Patterson C, Shah P;  
XX WPI: 2001-112727/12.  
DR P-PSDB; AAB60488.  
XX Human cell cycle and proliferation proteins and polynucleotides are  
PT used to treat, diagnose and prevent immune, developmental and cell  
PT signaling disorders and cell proliferative disorders including cancer -  
XX Claim 5; Page 190-191; 205pp; English.  
XX Sequences AAF59590-AAF59643 represent cDNAs encoding 54 human  
CC cell cycle and proliferation proteins (CCYPR), AAB60453-AAB60506.  
CC CCYPR and agonists of CCYPR are used to treat diseases or conditions  
CC associated with decreased expression of functional CCYPR, while CCYPR  
CC antagonists are used to treat diseases or conditions associated with  
CC overexpression of functional CCYPR. Monoclonal or polyclonal antibodies  
CC to CCYPR may be used in enzyme-linked immunosorbent assays (ELISA) or  
CC radioimmunoassays to detect CCYPR. CCYPR itself may be used to detect  
CC compounds e.g., antibodies, oligonucleotides and proteins (receptors)  
CC that specifically bind to CCYPR, and in drug screening methods to  
CC identify compounds that modulate the activity of CCYPR. CCYPR  
CC nucleotides can be used to generate transgenic animal models of human  
CC disease, and can be used in gene therapy in target cells with genetic  
CC abnormalities with respect to the expression of CCYPR for the  
CC treatment or prevention of a disorder associated with CCYPR.  
CC Diseases which can be diagnosed, treated and prevented using CCYPR  
CC proteins, nucleic acids, agonists or antagonists include immune,  
CC developmental and cell signalling disorders, and cell proliferative  
CC disorders including cancer. Specific examples of these disorders  
CC include anaemia, epilepsy, arteriosclerosis, asthma, cancer, allergies,  
CC diabetes mellitus, disorders of the menstrual cycle and infections  
CC caused by bacteria.  
XX SQ Sequence 2555 BP; 577 A; 703 C; 722 G; 547 T; 0 other;

Query Match 100.0%; Score 2555; DB 22; Length 2555;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTATGACAGTGGGAGGATGTGATGAAAGGATGAAGTGGAGTGC 60  
DB 1 AGCAGTCCCACTATGACAGTGGGAGGATGTGATGAAAGGATGAAGTGGAGTGC 60  
QY 61 TCAACAGTGTGCTGTGCTCCAGCCGGGTGTACTGATCGCTCTGTATCCAGACAG 120  
DB 61 TCAACAGTGTGCTGTGCTCCAGCCGGGTGTACTGATCGCTCTGTATCCAGACAG 120  
QY 121 CAGGATATCGGCTGTGCTTGGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 180  
DB 121 CAGGATATCGGCTGTGCTTGGTATGAAGGCTTTGAAATGACGCCAGCCATGACTTCT 180  
QY 181 GGTGCAACCTGGGAACAGTGTGATGTCACCCCATTTGGTGTGTCATCAACAGCAAGA 240  
DB 181 GGTGCAACCTGGGAACAGTGTGATGTCACCCCATTTGGTGTGTCATCAACAGCAAGA 240  
QY 241 TCTAGTGCCTCCAGGACCATCATGCTCAAGTTCACCGCTGGAAGGGTACCTCATGA 300  
DB 241 TCTAGTGCCTCCAGGACCATCATGCTCAAGTTCACCGCTGGAAGGGTACCTCATGA 300  
QY 301 AACGCTGGTGGCTTCCAGGACGCTTCCGCTGGATTTCCATCAAGATGGTGGAGACGA 360  
DB 301 AACGCTGGTGGCTTCCAGGACGCTTCCGCTGGATTTCCATCAAGATGGTGGAGACGA 360  
QY 361 TGAAGTACCCCTTTAGGCAGGACATGCGGCTGGAGTGGTGGAGTGGAGTGGTGCAC 420  
DB 361 TGAAGTACCCCTTTAGGCAGGACATGCGGCTGGAGTGGTGGAGTGGAGTGGTGCAC 420  
QY 421 GCACCTCGCATGGCTGTGGTGGACACAGTAATCGGGGGTTCGCTACGGCTCTCTACGAGG 480  
DB 421 GCACCTCGCATGGCTGTGGTGGACACAGTAATCGGGGGTTCGCTACGGCTCTCTACGAGG 480

QY 481 ATGGTGACAGTGCACGACGACTTCTGTGTCACATGTGGAGCCCCCTGATCCACCCAGTGG 540  
DB 481 ATGGTGACAGTGCACGACGACTTCTGTGTCACATGTGGAGCCCCCTGATCCACCCAGTGG 540  
QY 541 GTTGTGACAGTGTGGGCGACGCGATCAGATGTCAGAGGCGAAGTGCATGGGCC 600  
DB 541 GTTGTGACAGTGTGGGCGACGCGATCAGATGTCAGAGGCGAAGTGCATGGGCC 600  
QY 601 ATCACCACCACCTTCGGGAAGATCTACTGTGATGCGCTTCCCTTACCTCTTCAAGAGGTAC 660  
DB 601 ATCACCACCACCTTCGGGAAGATCTACTGTGATGCGCTTCCCTTACCTCTTCAAGAGGTAC 660  
QY 661 GAGCAGTCTACACAGAGCGGTTGGTTGAGGAAGGATGAAGTGGAGGCTATGACC 720  
DB 661 GAGCAGTCTACACAGAGCGGTTGGTTGAGGAAGGATGAAGTGGAGGCTATGACC 720  
QY 721 CCTGAATCTGGGCAACATCTGCGTGGCACTGTCTGTAAGTCTTCTCTGATGATACC 780  
DB 721 CCTGAATCTGGGCAACATCTGCGTGGCACTGTCTGTAAGTCTTCTCTGATGATACC 780  
QY 781 TGATGATCTGTGACGCGGGGCTCTCCACAGATGGCTTGGACTGCTTGTCTACCATG 840  
DB 781 TGATGATCTGTGACGCGGGGCTCTCCACAGATGGCTTGGACTGCTTGTCTACCATG 840  
QY 841 CCTCTTCCACGCGCATCTTCCGCGCATCTTCTGTGCAAGATGACATTTGAGCTCACAC 900  
DB 841 CCTCTTCCACGCGCATCTTCCGCGCATCTTCTGTGCAAGATGACATTTGAGCTCACAC 900  
QY 901 CGCCAAAAGGTTATGAGGACACAGATCTTCACTGGGAGAACTACTTTGGAGAGACCAAGT 960  
DB 901 CGCCAAAAGGTTATGAGGACACAGATCTTCACTGGGAGAACTACTTTGGAGAGACCAAGT 960  
QY 961 CGAAAGCGCTCCATCGAGACTCTTTAATCATGGATTGCGCAACCATGGCTTCAAGGTGG 1020  
DB 961 CGAAAGCGCTCCATCGAGACTCTTTAATCATGGATTGCGCAACCATGGCTTCAAGGTGG 1020  
QY 1021 GCATGAAGCTGGAGCGCTGGACCTGATGGAGCCCCGGCTCATCTGTGTGGCCACGGTGA 1080  
DB 1021 GCATGAAGCTGGAGCGCTGGACCTGATGGAGCCCCGGCTCATCTGTGTGGCCACGGTGA 1080  
QY 1081 AACAGTGTGTGATCGGCTCTCTCAGCATCCATTTGACGGCTGGGACAGCGATACGACC 1140  
DB 1081 AACAGTGTGTGATCGGCTCTCTCAGCATCCATTTGACGGCTGGGACAGCGATACGACC 1140  
QY 1141 AGTGGGTGGACTGCGAGTCCCAGACATCTACCCGTCGGCTGGTGTGAGCTCACCGCT 1200  
DB 1141 AGTGGGTGGACTGCGAGTCCCAGACATCTACCCGTCGGCTGGTGTGAGCTCACCGCT 1200  
QY 1201 ACCAGCTCCAGCTCTCTGTGGCGCAGAACCGGCCACACCCGCTGAAGGCCAAAGAGGCCA 1260  
DB 1201 ACCAGCTCCAGCTCTCTGTGGCGCAGAACCGGCCACACCCGCTGAAGGCCAAAGAGGCCA 1260  
QY 1261 CAAAGAGAAAGAAACAGTTTGGGAGAAAGGAAAGAAATCCGCGCTACGAGCGC 1320  
DB 1261 CAAAGAGAAAGAAACAGTTTGGGAGAAAGGAAAGAAATCCGCGCTACGAGCGC 1320  
QY 1321 GACCCCTCAGACAGAGGCTCCAAAGAGCCCTGCTGGAGGACGACCTCAGGGTGCAGGA 1380  
DB 1321 GACCCCTCAGACAGAGGCTCCAAAGAGCCCTGCTGGAGGACGACCTCAGGGTGCAGGA 1380  
QY 1381 AGATCTCGTGGAGCTGTCTCTGCGGAGATCATTTGCTGTGCGTGAAGAGAGCATC 1440  
DB 1381 AGATCTCGTGGAGCTGTCTCTGCGGAGATCATTTGCTGTGCGTGAAGAGAGCATC 1440  
QY 1441 TAGAGTGGCTTCGCGGACAGAGGCTTCAAGTCCAGAGCTCGCTGTCTCCGTCGAGAAC 1500  
DB 1441 TAGAGTGGCTTCGCGGACAGAGGCTTCAAGTCCAGAGCTCGCTGTCTCCGTCGAGAAC 1500  
QY 1501 TCAAGCAGAAACAGACGACTGAGCTTCTCTGCTCCAGCTGGCTTCTAGCTGGAAGCC 1560  
DB 1501 TCAAGCAGAAACAGACGACTGAGCTTCTCTGCTCCAGCTGGCTTCTAGCTGGAAGCC 1560

AA158732;  
22-OCT-2001 (first entry)  
Human polynucleotide SEQ ID NO 935.  
Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.  
Homo sapiens.  
WO200153312-A1.  
26-JUL-2001.  
26-DEC-2000; 2000WO-US34263.  
21-JAN-2000; 2000US-0488725.  
25-APR-2000; 2000US-0552317.  
09-JUL-2000; 2000US-0598042.  
19-JUL-2000; 2000US-0620312.  
03-AUG-2000; 2000US-0653450.  
14-SEP-2000; 2000US-0662191.  
19-OCT-2000; 2000US-0693036.  
29-NOV-2000; 2000US-0727344.  
(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue A, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
WPI; 2001-442253//47.  
P-PSDB; AAM39576.  
Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries -  
Claim 1; SEQ ID NO 935; 10078pp; English.  
The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAM42213) with nontropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.  
Note: The sequence data for this patent did not form part of the printed specification.  
Sequence 2781 BP; 632 A; 776 C; 784 G; 589 T; 0 other;  
Query Match 99.8%; Score 2550.8; DB 22; Length 2781;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 2552; Conservative 0; Mismatches 2; Indels 0; Gaps 0  
QY 1 AGCAGTCCCACTCTATGACCATGTGGAGAGTGTGATGAAGCGATGAAGTGGAGTGC 60  
DB 228 AGCAGTCCCACTCTATGACCATGTGGAGAGTGTGATGAAGCGATGAAGTGGAGTGC 287  
QY 61 TCACAGTGTGTGTCTCTCCCGAGCGGGTGTACTGGATCGCCTCTGTCACTCCAGACAG 120



288 TCAACAGTGTGCTGCTGCCAGCCGGGTGACTGGATCGGCTCTGTCTATCCAGACAG 347  
121 CAGGGTATCGGGTCTCTCTTGGTATGAAGGCTTTGAAATAGCGCCAGCCATGACTTCT 180  
348 CAGGGTATCGGGTCTCTCTTGGTATGAAGGCTTTGAAATAGCGCCAGCCATGACTTCT 407  
181 GGTCAACCTTGGGAACAGTGGATGTCACCCCATTTGGCTGGTGTGCATCAACAGCAAGA 240  
408 GGTCAACCTTGGGAACAGTGGATGTCACCCCATTTGGCTGGTGTGCATCAACAGCAAGA 467  
241 TCTTAGTCCGCCACGACGACCATCCATGCCAAGTTTCAACCGACTGGAAGGCTTACTCTATCA 300  
468 TCTTAGTCCGCCACGACGACCATCCATGCCAAGTTTCAACCGACTGGAAGGCTTACTCTATCA 527  
301 AACGGCTGGTGGGCTTCAGAGACGTTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 360  
528 AACGGCTGGTGGGCTTCAGAGACGTTTCCCGTGGATTTCCACATCAAGATGGTGGAGAGCA 587  
361 TGAAGTACCCCTTTAGCAGCGCATCGGCTGCAAGTGTGGAAGTGGGAGGCGAAGTGCATGGGCC 600  
588 TGAAGTACCCCTTTAGCAGCGCATCGGCTGCAAGTGTGGAAGTGGGAGGCGAAGTGCATGGGCC 647  
421 GCATCGCATGGCTGTGGTGGACACAGTAAATCGGGGGTGCCTTACCGCTCTTACGAGG 480  
648 GCATCGCATGGCTGTGGTGGACACAGTAAATCGGGGGTGCCTTACCGCTCTTACGAGG 707  
481 ATGGTGACAGTACGACGACTTCTGTGGCCACATGTGGAGCCCGCTGATCCACCCAGTGG 540  
708 ATGGTGACAGTACGACGACTTCTGTGGCCACATGTGGAGCCCGCTGATCCACCCAGTGG 767  
541 GTTGGTCACGACGTGTGGGCCACGCGATCAAGATGTACAGAGAGGCGAAGTGCATGGGCC 600  
768 GTTGGTCACGACGTGTGGGCCACGCGATCAAGATGTACAGAGAGGCGAAGTGCATGGGCC 827  
601 ATCACCCACCTTCCGAGAGTACTGTGATGCCCTTCTTACCTTCTTCAAGAGGTTAC 660  
828 ATCACCCACCTTCCGAGAGTACTGTGATGCCCTTCTTACCTTCTTCAAGAGGTTAC 887  
661 GAGCAGTCTACACAGAGGCGGTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 720  
888 GAGCAGTCTACACAGAGGCGGTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 947  
721 CCTGATCTGGGCAACATCTGGTGGCACTGTCTGTAAGTGTCTCTGGATGATACC 780  
948 CCTGATCTGGGCAACATCTGGTGGCACTGTCTGTAAGTGTCTCTGGATGATACC 1007  
781 TGAATGATCTGTGGAGCGGGGCCCTCCACAGATGGCTTGGACTGGTCTGCTACCATG 840  
1008 TGAATGATCTGTGGAGCGGGGCCCTCCACAGATGGCTTGGACTGGTCTGCTACCATG 1067  
841 CTTCTTCCACGCGCATCTTCCCGGCCACCTTCTGTACAGAGATGACATTTGAGCTTCACAC 900  
1068 CTTCTTCCACGCGCATCTTCCCGGCCACCTTCTGTACAGAGATGACATTTGAGCTTCACAC 1127  
901 CGCCAAAGGTTATGAGGCACAGACTTTCATCTGGAGAGACTTCTTGGAGAGACCAAGT 960  
1128 CGCCAAAGGTTATGAGGCACAGACTTTCATCTGGAGAGACTTCTTGGAGAGACCAAGT 1187  
961 CGAAAGCCGCTCCATCGAGACTCTTTAAATGATGATTTGCCAAACCATGGCTTCAAGGTGG 1020  
1188 CGAAAGCCGCTCCATCGAGACTCTTTAAATGATGATTTGCCAAACCATGGCTTCAAGGTGG 1247  
1021 GCATGAGTCTGAGGCCGTTGACCTTGTAGAGCCCGGCTCATCTGTGTGCCACGTTGA 1080  
1248 GCATGAGTCTGAGGCCGTTGACCTTGTAGAGCCCGGCTCATCTGTGTGCCACGTTGA 1307  
1081 AACGAGTGTGATCGGCTCTTCAGCATCCATTTTACGGCTGGACAGCGAGTACGACC 1140  
1308 AACGAGTGTGATCGGCTCTTCAGCATCCATTTTACGGCTGGACAGCGAGTACGACC 1367  
1141 AGTGGGTGATCTGAGTCCCGCAGACATCTACCCCGTGGTGGTGTGAGTCTACCCGCT 1200  
1368 AGTGGGTGATCTGAGTCCCGCAGACATCTACCCCGTGGTGGTGTGAGTCTACCCGCT 1427

1201 ACCAGTCTCCAGCTCTCTGTGGCCGACGCGCCACACCGCTGAAGGCCCAAGAGGCCA 1260  
1428 ACCAGTCTCCAGCTCTCTGTGGCCGACGCGCCACACCGCTGAAGGCCCAAGAGGCCA 1487  
1261 CAAAGAGAAAAGAAAACAGTTTGGGAAGAAAAGAAAAGAAATCCCGCCCACTAAGACGC 1320  
1488 CAAAGAGAAAAGAAAACAGTTTGGGAAGAAAAGAAAAGAAATCCCGCCCACTAAGACGC 1547  
1321 GACCCCTCAGACAGGGTCCAGAGAGCCCTGTCTGGAGAGAGACCCCTCAGGGTCCAGGA 1380  
1548 GACCCCTCAGACAGGGTCCAGAGAGCCCTGTCTGGAGAGAGACCCCTCAGGGTCCAGGA 1607  
1381 AGATCTCGTCCGAGACCTGTTCTCTGGCCAGATCATTTGCTGTGCGGTGTGAAGAGAGCATC 1440  
1608 AGATCTCGTCCGAGACCTGTTCTCTGGCCAGATCATTTGCTGTGCGGTGTGAAGAGAGCATC 1667  
1441 TAGACGTGGCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTGTCTCGTCCAGAAACA 1500  
1668 TAGACGTGGCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTGTCTCGTCCAGAAACA 1727  
1501 TCAAGCAGGAAAACAGACGACTGAGCCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGCC 1560  
1728 TCAAGCAGGAAAACAGACGACTGAGCCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGCC 1787  
1561 AGCCAGCGTTTCTCTACCAACCAACCATGCTCCACTTCCACTGACTTTGGCTTGGAGACTGA 1620  
1788 AGCCAGCGTTTCTCTACCAACCAACCATGCTCCACTTCCACTGACTTTGGCTTGGAGACTGA 1847  
1621 TCCTCTCTGTGTAATTTCTGCCCGGTGCTGTGAAGCTGGACGGTGGAGAGACCTGCTGGG 1680  
1848 TCCTCTCTGTGTAATTTCTGCCCGGTGCTGTGAAGCTGGACGGTGGAGAGACCTGCTGGG 1907  
1681 GTCTCTGGACCCCGCTGTTGCTTCTGCCCTCCCTCTGGAAGGTCTATATGACGGGC 1740  
1908 GTCTCTGGACCCCGCTGTTGCTTCTGCCCTCCCTCTGGAAGGTCTATATGACGGGC 1967  
1741 CGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCAGAGTTTCCCAAAGCTG 1800  
1968 CGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCAGAGTTTCCCAAAGCTG 2027  
1801 GAACGCTAGTGGCTGCTTCTTCTTAAAGATGGCTCCCGCCAGCCCGCCAGCCCTCAG 1860  
2028 GAACGCTAGTGGCTGCTTCTTCTTAAAGATGGCTCCCGCCAGCCCGCCAGCCCTCAG 2087  
1861 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAATCTGCTG 1920  
2088 TTGCCAGGGATGGGGCCACCACTGTCACTGTGGAATACAAGACAGTGAATCTGCTG 2147  
1921 CCTGAACGAGTCAATGTAATTAAGTTCTAGAGCAGCTCTCTGAGCAGGATTAAGTCCCT 1980  
2148 CCTGAACGAGTCAATGTAATTAAGTTCTAGAGCAGCTCTCTGAGCAGGATTAAGTCCCT 2207  
1981 GACAGTGAAGTGTGGTGGGGCAGCCTCTGCTCAAAAATTCACCAAGCAGAAATGSCCT 2040  
2208 GACAGTGAAGTGTGGTGGGGCAGCCTCTGCTCAAAAATTCACCAAGCAGAAATGSCCT 2267  
2041 CTCAGCCTCATGTGTTGTTCT 2100  
2268 CTCAGCCTCATGTGTTGTTCT 2327  
2101 CTTGTCTCCGAGCTAAGAGCAGTGAACAGGATGTGAATTTTGGCGACTGTGTGTGG 2160  
2328 CTTGTCTCCGAGCTAAGAGCAGTGAACAGGATGTGAATTTTGGCGACTGTGTGTGG 2387  
2161 CTTGTAGCTGCTTCTGTGTTTGTGAGGACTGACTCCCATTTTCTTAAAGAAATGCCCCC 2220  
2388 CTTGTAGCTGCTTCTGTGTTTGTGAGGACTGACTCCCATTTTCTTAAAGAAATGCCCCC 2447  
2221 GGGAGGACATTTGGGAGGAGATGGCTGAGTGTGACATTTTGGCTCTGTCTCTCTCTCTCT 2280  
2448 GGGAGGACATTTGGGAGGAGATGGCTGAGTGTGACATTTTGGCTCTGTCTCTCTCTCTCTCT 2507



QY 2281 GAAGCCCGCTAAATATTCATCAAGATTCTTTGAGTAAAGGTCACCTTCTGA 2340  
 DB 2508 GAAGCCCGCTAAATATTCATCAAGATTCTTTGAGTAAAGGTCACCTTCTGA 2567  
 QY 2341 CTGAGCCTCTAGAGAGCTGGGCTGTATGTTCTTTGGCTTTTGTCTTACCTAAATG 2400  
 DB 2568 TTGAGCCTCTAGAGAGCTGGGCTGTATGTTCTTTGGCTTTTGTCTTACCTAAATG 2627  
 QY 2401 AAGAAACCATCCCTGGAGGGCCCTGACACAGAACCTTCAGACAGAGATGACAGAGCT 2460  
 DB 2628 AAGAAACCATCCCTGGAGGGCCCTGACACAGAACCTTCAGACAGAGATGACAGAGCT 2687  
 QY 2461 GGAGGACACATCTAGCTGCCATTGCAACCTCATCTGGGCTCCCGAGACTCTGTGTGTGAGA 2520  
 DB 2688 GGAGGACACATCTAGCTGCCATTGCAACCTCATCTGGGCTCCCGAGACTCTGTGTGTGAGA 2747  
 QY 2521 AATTAACCCCTCTGCTTGTGAGAAAAA 2554  
 DB 2748 AATTAACCCCTCTGCTTGTGAGAAAAA 2781

RESULT 3

ACC50583  
 ID ACC50583 standard; cDNA; 3530 BP.

XX ACC50583;

XX 12-JUN-2003 (first entry)

XX Human secreted protein coding sequence, SEQ ID 250.

XX Cardiant; antiarrhythmic; antiarteriosclerotic; vasotropic; cytostatic;  
 XX vulnerary; antiinflammatory; nootropic; neuroprotective;  
 KW antiparkinsonian; gene therapy; human; cardiovascular disorder;  
 KW gene; ss.

XX Homo sapiens.

XX WQ200295010-A2.

XX 28-NOV-2002.

XX 19-MAR-2002; 2002WO-US09785.

XX 21-MAR-2001; 2001US-277340P.

XX 19-JUL-2001; 2001US-306171P.

XX 13-NOV-2001; 2001US-331287P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SW;

XX WPI; 2003-129429/12.

XX Novel human secreted proteins, useful for detecting, preventing,  
 PT diagnosing, prognosticating, treating and/or ameliorating  
 PT cardiovascular disorders such as arrhythmia -

PS Claim 21; SEQ ID 250; 1881pp; English.

XX The present invention relates to novel human secreted proteins  
 CC (AB47633-AB48145) and their coding sequences (ACC50344-ACC50856). The  
 CC proteins and their coding sequences are useful for the preparation of a  
 CC diagnostic or pharmaceutical composition for diagnosing or treating a  
 CC cardiovascular disorder (e.g., arrhythmia, tachycardia, cardiac arrest,  
 CC coronary arteriosclerosis and myocardial ischemia), neural disorders,  
 CC immune system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, renal disorders,  
 CC wound healing and epithelial cell proliferation, to treat inflammation or  
 CC infection, for treating thrombosis and arteriosclerosis, for treating or  
 CC preventing neural damage which occurs in neuronal disorders or  
 CC neurodegenerative conditions such as Alzheimer's disease and Parkinson's

CC disease, to enhance bone and periodontal regeneration and aid in tissue  
 CC transplants or bone grafts, to prevent skin aging or hair loss, to  
 CC stimulate growth and differentiation of haematopoietic cells and bone  
 CC marrow cells when used in combination with other cytokines, to maintain  
 CC organs before transplantation or for supporting cell culture of primary  
 CC tissues, to increase or decrease differentiation or proliferation of  
 CC embryonic stem cells, or to modulate mammalian characteristics or  
 CC metabolism.  
 CC Note: The sequence data for this patent was published in electronic  
 CC format and is available from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3530 BP; 815 A; 945 C; 972 G; 790 T; 8 other;

Query Match 99.6%; Score 2544; DB 25; Length 3530;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2544; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTCTATGACCACTGGGAGGATGATGAAAGGGATGAAGTGGAGTGC 60  
 DB 895 AGCAGTCCCACTCTATGACCACTGGGAGGATGATGAAAGGGATGAAGTGGAGTGC 954  
 QY 61 TCACAGTGTATCTGTCTCCCAAGCGGGTGTACTTGGATCGCTCTGTCTATCCAGACAG 120  
 DB 955 TCACAGTGTATCTGTCTCCCAAGCGGGTGTACTTGGATCGCTCTGTCTATCCAGACAG 1014  
 QY 121 CAGGATATCGGGTGTCTGCTTCGGTATGAAGCTTTGAAATGACGCCAGCCATGACTTCT 180  
 DB 1015 CAGGATATCGGGTGTCTGCTTCGGTATGAAGCTTTGAAATGACGCCAGCCATGACTTCT 1074  
 QY 181 GGTCAACCTCGGAAACAGTGTGATGTCACCCCATGCTGCTGTCATCAACAGCAGA 240  
 DB 1075 GGTCAACCTCGGAAACAGTGTGATGTCACCCCATGCTGCTGTCATCAACAGCAGA 1134  
 QY 241 TCCTAGTGTCCCAAGCAACCATCCATGCCAAGTTCACCGACTGGAAGGGTACCTCATGA 300  
 DB 1135 TCCTAGTGTCCCAAGCAACCATCCATGCCAAGTTCACCGACTGGAAGGGTACCTCATGA 1194  
 QY 301 AACGGCTGTGGGCTCCAGAGCGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 360  
 DB 1195 AACGGCTGTGGGCTCCAGAGCGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 1254  
 QY 361 TGAAGTACCCCTTTAGCAGCGGATGCGGTGGAAGTGTGGACAAGTCCCAAGTGTAC 420  
 DB 1255 TGAAGTACCCCTTTAGCAGCGGATGCGGTGGAAGTGTGGACAAGTCCCAAGTGTAC 1314  
 QY 421 GCATCGCATGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTACGGCTCTCTACGAGG 480  
 DB 1315 GCATCGCATGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTACGGCTCTCTACGAGG 1374  
 QY 481 ATGGTGACAGTGACGACGACTTCTGGTGCCACATGTGGAGCCCGCTGATCCACCGAGTGG 540  
 DB 1375 ATGGTGACAGTGACGACGACTTCTGGTGCCACATGTGGAGCCCGCTGATCCACCGAGTGG 1434  
 QY 541 GTTGGTCAACAGTGTGGGCCACCGCATCAAGATGTGACAGAGGGGAGATGACATGGCCCC 600  
 DB 1435 GTTGGTCAACAGTGTGGGCCACCGCATCAAGATGTGACAGAGGGGAGATGACATGGCCCC 1494  
 QY 601 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCGCTTCTTACCTCTTCAAGAAGTAC 660  
 DB 1495 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCGCTTCTTACCTCTTCAAGAAGTAC 1554  
 QY 661 GAGCAGTCTACACAGAAGCGGTGGTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 720  
 DB 1555 GAGCAGTCTACACAGAAGCGGTGGTGGTTTGGAGAGGGATGAAGTGGAGGCCATTGACC 1614  
 QY 721 CCTGAATCTGGGACACATCTCGTGGCACTGTCTGTAAAGTTCCTCTGGATGGATACC 780  
 DB 1615 CCTGAATCTGGGACACATCTCGTGGCACTGTCTGTAAAGTTCCTCTGGATGGATACC 1674  
 QY 781 TGATGATCTGTGGACGGGGGGCGCTCCACAGATGGCTTGGACTGGTTCCTCTACCATG 840  
 DB 1675 TGATGATCTGTGGACGGGGGGCGCTCCACAGATGGCTTGGACTGGTTCCTCTACCATG 1734

841 CCTTTCCCAAGCCATCTTCCCGGCGACCTTCTGTGTCAGAGAAATGATGAGCTCACAC 900  
1735 CCTTTCCCAAGCCATCTTCCCGGCGACCTTCTGTGTCAGAGAAATGATGAGCTCACAC 1794  
901 CGCCAAAAGGTTATGAGGACACAGACTTTCACCTGGGAGAGACTTTCGAGAGACCAAGT 960  
1795 CGCCAAAAGGTTATGAGGACACAGACTTTCACCTGGGAGAGACTTTCGAGAGACCAAGT 1854  
961 CGAAAGCCGCTCCATCGAGACTCTTTAACTAATGATTGCCCAACCACTGCTTCAAGGTGG 1020  
1855 CGAAAGCCGCTCCATCGAGACTCTTTAACTAATGATTGCCCAACCACTGCTTCAAGGTGG 1914  
1021 GCATGAAGCTGGAGGCGCTGAGCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080  
1915 GCATGAAGCTGGAGGCGCTGAGCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1974  
1081 AACGAGTGTGTCATCGGCTCTCTCAGCATCCACATTTGACGCGCTGGGACACGAGTACGACC 1140  
1975 AACGAGTGTGTCATCGGCTCTCTCAGCATCCACATTTGACGCGCTGGGACACGAGTACGACC 2034  
1141 AGTGGGTGGAGTGGAGTCCCGAGACATCTACCCCGTGGCTGGTGTGAGCTCACCGGCT 1200  
2035 AGTGGGTGGAGTGGAGTCCCGAGACATCTACCCCGTGGCTGGTGTGAGCTCACCGGCT 2094  
1201 ACCAGCTCCAGGCTCTGTGGCGCGCAGAACCGGCGCACACCGCTGAAGGCCAAAGAGGCCA 1260  
2095 ACCAGCTCCAGGCTCTGTGGCGCGCAGAACCGGCGCACACCGCTGAAGGCCAAAGAGGCCA 2154  
1261 CAAAGAAAGAAAGAAACAGTTTGGGAAAGAAAGAAAGAAAGTCCCGCCACTAAGACGC 1320  
2155 CAAAGAAAGAAAGAAACAGTTTGGGAAAGAAAGAAAGAAAGTCCCGCCACTAAGACGC 2214  
1321 GACCCCTCAGACAGAGGCTCCAGAGAGCCCTCTCGAGAGCAGACCTCAGGGTGCAGGA 1380  
2215 GACCCCTCAGACAGAGGCTCCAGAGAGCCCTCTCGAGAGCAGACCTCAGGGTGCAGGA 2274  
1381 AGATCTCGTGGAGCTGTCTCTGGCGAGATCAATTCCTGTGTGAGGAAAGAGCATC 1440  
2275 AGATCTCGTGGAGCTGTCTCTGGCGAGATCAATTCCTGTGTGAGGAAAGAGCATC 2334  
1441 TAGAGTGGCTCGCCGACAGAGCTTCAAGTCCAGAGTGGCTGTCTCCGTGAGAAC 1500  
2335 TAGAGTGGCTCGCCGACAGAGCTTCAAGTCCAGAGTGGCTGTCTCCGTGAGAAC 2394  
1501 TCAAGCAGAAACAGACGACTGAGCTTCTCGCTCCAGGCTGGCTTCTAGCTGAGAGCC 1560  
2395 TCAAGCAGAAACAGACGACTGAGCTTCTCGCTCCAGGCTGGCTTCTAGCTGAGAGCC 2454  
1561 AGCCAGAGCTTCTCTACACACACACACGAGCTCCACCTGACTTTGGCTTGGAGACTGA 1620  
2455 AGCCAGAGCTTCTCTACACACACACACGAGCTCCACCTGACTTTGGCTTGGAGACTGA 2514  
1621 TCCTCTCTGTAAATTCCTCGCGTGTGTGAAGGCTGGACGCTGGAGGAGCTGCTGGG 1680  
2515 TCCTCTCTGTAAATTCCTCGCGTGTGTGAAGGCTGGACGCTGGAGGAGCTGCTGGG 2574  
1681 GTCTCTGGACCCGCTGTCTCTCGCTTCTCGCTTCCCTGTGAAAGGCTATATGACGCGC 1740  
2575 GTCTCTGGACCCGCTGTCTCTCGCTTCTCGCTTCCCTGTGAAAGGCTATATGACGCGC 2634  
1741 CGCCTTGGGCCCCAGACTCGTGTGAAACCACTTTTCAGCCAGAGTTCGCCAAGCTG 1800  
2635 CGCCTTGGGCCCCAGACTCGTGTGAAACCACTTTTCAGCCAGAGTTCGCCAAGCTG 2694  
1801 GAACCTAGTCTGCTCTCTCTTAAAGATGGCTTCCCGCCGACCCGCGACGCTCAG 1860  
2695 GAACCTAGTCTGCTCTCTCTTAAAGATGGCTTCCCGCCGACCCGCGACGCTCAG 2754  
1861 TTGCCAGGATGGGGCCACCACTGTACACTGTGGAAATACAGACAGTGAATCTGTCTG 1920  
2755 TTGCCAGGATGGGGCCACCACTGTACACTGTGGAAATACAGACAGTGAATCTGTCTG 2814

1921 CCTGAACGAGTCATGTAATTAAGTCTTAGACAGAGCTCTCTGAGCAGGATAAGGTCCCT 1980  
2815 CCTGAACGAGTCATGTAATTAAGTCTTAGACAGAGCTCTCTGAGCAGGATAAGGTCCCT 2874  
1981 GACAGTGAAGTGTGTGTGGGGGAGGCTCTGCTTCAAAAATTCACCAAGCAGAAATGCT 2040  
2875 GACAGTGAAGTGTGTGTGGGGGAGGCTCTGCTTCAAAAATTCACCAAGCAGAAATGCT 2934  
2041 CTCAGCTCATGTCTTGGTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGAGCCAG 2100  
2935 CTCAGCTCATGTCTTGGTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGAGCCAG 2994  
2101 CTGTCTCGGAGCTTAAGAGCAGTGAACAGATGTGGATTTTGGGCACTGTGTGTGG 2160  
2995 CTGTCTCGGAGCTTAAGAGCAGTGAACAGATGTGGATTTTGGGCACTGTGTGTGG 3054  
2161 CTTTGAAGTCTTCTGTCTGTGTGTGGAGTGAACAGATGTGGATTTTGGGCACTGTGTGTGG 3114  
3055 CTTTGAAGTCTTCTGTCTGTGTGTGGAGTGAACAGATGTGGATTTTGGGCACTGTGTGTGG 3174  
2221 GGGGAGCAGATTTGGGAGAGATGGGCTGTGATGTGCACTTTGGCTCTGTCTTACCTGTCT 2280  
3115 GGGGAGCAGATTTGGGAGAGATGGGCTGTGATGTGCACTTTGGCTCTGTCTTACCTGTCT 3174  
2281 GAAGCCCGCTTAAATAATTCATCCAGATTCCTTTGTAGTTTAAAGGTCAGTTCTGA 2340  
3175 GAAGCCCGCTTAAATAATTCATCCAGATTCCTTTGTAGTTTAAAGGTCAGTTCTGA 3234  
2341 CTGAGGCTCTAGAGAGTGGGCTTGTATGTTCTTTGGCTTTTGTCTTACCTAAATG 2400  
3235 CTGAGGCTCTAGAGAGTGGGCTTGTATGTTCTTTGGCTTTTGTCTTACCTAAATG 3294  
2401 AAGAAACCATGCTTGGAGGGCGGTGAACACAGAACCTTCAAGACAGATGACAGCT 2460  
3295 AAGAAACCATGCTTGGAGGGCGGTGAACACAGAACCTTCAAGACAGATGACAGCT 3354  
2461 GGAGGACACATCTAGCTGCAATTCACCTCACTGGGCTCCCGAGACTCTGTGTGTGAGA 2520  
3355 GGAGGACACATCTAGCTGCAATTCACCTCACTGGGCTCCCGAGACTCTGTGTGTGAGA 3414  
2521 AATTAAACCCCTGCTTGTGTGAG 2544  
3415 AATTAAACCCCTGCTTGTGTGAG 3438

## RESULT 4

ABZ71320  
ID ABZ71320 standard; cDNA; 3530 BP.

XX AC ABZ71320;

XX AC ABZ71320;

DT 04-APR-2003 (first entry)

XX Secreted protein-encoding gene 131 cDNA clone HOBJ92, SEQ ID NO:141.

XX Human; secreted protein; digestive disorder; gastrointestinal disorder;  
XX mouth; esophagus; stomach; small intestine; large intestine; liver;  
XX biliary tract; pancreas; cancer; tumour; hyperproliferative disorder;  
XX immune disorder; inflammation; infection; wound healing; drug screening;  
XX chromosome identification; chromosome mapping; cytostatic;  
XX anti-inflammatory; immunosuppressive; vulnery; gene therapy; gene; ss.

OS Homo sapiens.

XX WO200276488-A1.

PN 03-OCT-2002.

XX 19-MAR-2002; 2002WO-US08276.

XX 21-MAR-2001; 2001US-277340P.

PR 19-JUL-2001; 2001US-306171P.

PR 13-NOV-2001; 2001US-331287P.





Db 252 GTTGCACTGGGAAACAGTGGATGTCCACCCCATTTGGCTGGTGCATCAACAGCAAGA 311  
Qy 241 TCTTAGTGCCCCCAACGACCATCATGCCAAGTTACCGACTGGAAGGGCTACTCATGA 300  
Db 312 TCTTAGTGCCCCCAACGACCATCATGCCAAGTTACCGACTGGAAGGGCTACTCATGA 371  
Qy 301 AACGGCTGGTGGGCTCCAGACGCTTCCCGTGAGATTCACATCAAGATGGTGAGAGCA 360  
Db 372 AACGGCTGGTGGGCTCCAGACGCTTCCCGTGAGATTCACATCAAGATGGTGAGAGCA 431  
Qy 361 TGAAGTACCCCTTTAGSCAGGGCATCGGCTGGAAGTGGTGAAGTCCCAAGTCCCAAGTGTAC 420  
Db 432 TGAAGTACCCCTTTAGSCAGGGCATCGGCTGGAAGTGGTGAAGTCCCAAGTGTAC 491  
Qy 421 GCACCTGCATGGCTGGTGGACACAGTAATCGGGGGTGCCTACGGCTCCTCTACGAGG 480  
Db 492 GCACCTGCATGGCTGGTGGACACAGTAATCGGGGGTGCCTACGGCTCCTCTACGAGG 551  
Qy 481 ATGGTGACAGTGACGACGACTTCTGGTGGCCACATGTGGAGGCCCTGATCCACCCAGTGG 540  
Db 552 ATGGTGACAGTGACGACGACTTCTGGTGGCCACATGTGGAGGCCCTGATCCACCCAGTGG 611  
Qy 541 GTTGGTCAACAAGTGGGACAGGCATCAAGATGTACAGAGGGCAAGTGCATGGCCC 600  
Db 612 GTTGGTCAACAAGTGGGACAGGCATCAAGATGTACAGAGGGCAAGTGCATGGCCC 671  
Qy 601 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCCCTTACCTTTCAGAGAGGTAC 660  
Db 672 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCCCTTACCTTTCAGAGAGGTAC 731  
Qy 661 GAGCAGTCTACAGAGGCGGTGGTTTGGAGAGGGATGAAGCTGGAGGCCATTTGACC 720  
Db 732 GAGCAGTCTACAGAGGCGGTGGTTTGGAGAGGGATGAAGCTGGAGGCCATTTGACC 791  
Qy 721 CCTGTAATCTGGGACACATCTGGTGGCAACTGTCTGAAGTTCCTCTGGATGATACC 780  
Db 792 CCTGTAATCTGGGACACATCTGGTGGCAACTGTCTGAAGTTCCTCTGGATGATACC 851  
Qy 781 TGATGATCTGTGGACGGGGGCGCTCCACAGATGGCTTGACATGGTTCGTCTACCATG 840  
Db 852 TGATGATCTGTGGACGGGGGCGCTCCACAGATGGCTTGACATGGTTCGTCTACCATG 911  
Qy 841 CCTCTTCCACGCACTTCTCCGGGCACTTCTGTGAGAGGAATGACATTTGAGCTCACAC 900  
Db 912 CCTCTTCCACGCACTTCTCCGGGCACTTCTGTGAGAGGAATGACATTTGAGCTCACAC 971  
Qy 901 CGCCAAAGGTTATGAGGCACAGACTTTCACATGGGAGAACTACTTGGAGAGCAAGT 960  
Db 972 CGCCAAAGGTTATGAGGCACAGACTTTCACATGGGAGAACTACTTGGAGAGCAAGT 1031  
Qy 961 CGAAAGCCGCTCCATCGAGACTCTTTAATGGAATTTGCCAAACCATGGCTTCAAGGTGG 1020  
Db 1032 CGAAAGCCG-TCCATCGAGACTCTTTAATGGAATTTGCCAAACCATGGCTTCAAGGTGG 1090  
Qy 1021 GCATGAGCTGAGGCGCTGGACCTGATGAGGAGCCCGGCTCATCTGTGGCCACGGTGA 1080  
Db 1091 GCATGAGCTGAGGCGCTGGACCTGATGAGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1150  
Qy 1081 AACGAGTGGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140  
Db 1151 AACGAGTGGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1210  
Qy 1141 AGTGGTGCATCGAGTCCCGCAGACATCTACCCGCT-CGGCTGGTGTGAGCTCACCGGC 1199  
Db 1211 AGTGGTGCATCGAGTCCCGCAGACATCTACCCGCTGACGAGTGTGAGCTCACCGGC 1270  
Qy 1200 TACAGCTCCAGCTTCCGTGGCCGAGAACCGGCCACACGGCTGAAGGCCAAGAGGCC 1259  
Db 1271 TACAGCTCCAGCTTCCGTGGCCGAGAACCGGCCACACGGCTGAAGGCCAAGAGGCC 1330  
Qy 1260 ACAAGAGAAAAAAGAAACAGTTTGGGAGAAAAAGAAAGATCCCGCCCACTAAGACG 1319

Db 1331 ACAAGAGAAAAAAGAAACAGTTTGGGAGAAAAAGAAAGATCCCGCCCACTAAGACG 1390  
Qy 1320 CGACCCCTCAGACAGGGGTCCAAAGAGCCCTGTGGAGGACGACCCCTCAGGGTGCCAGG 1379  
Db 1391 CGACCCCTCAGACAGGGGTCCAAAGAGCCCTGTGGAGGACGACCCCTCAGGGTGCCAGG 1450  
Qy 1380 AAGATCTCGTGGAGCCCTGTTCTTGGCGAGATCAATGCTGTGCGTGTGAAGGAAGACAT 1439  
Db 1451 AAGATCTCGTGGAGCCCTGTTCTTGGCGAGATCAATGCTGTGCGTGTGAAGGAAGACAT 1510  
Qy 1440 CTAGACGTGGGCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTGCTCCGTCGAGAAC 1499  
Db 1511 CTAGACGTGGGCTCGCCCGACAGGCTTCAAGTCCAGAGCTGCCTGCTCTCGTGGAGAAC 1570  
Qy 1500 ATCAAGCAGGAAAACAGACGACTGAGCCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGC 1559  
Db 1571 ATCAAGCAGGAAAACAGACGACTGAGCCCTTCTGCTCCAGCCTGGCTTCTAGCTGGAGC 1630  
Qy 1560 CAGCCACGCTTCTCTACCAACACACCATGCTCCACCTGACTTTGGCTTGGAGACTG 1619  
Db 1631 CAGCCACGCTTCTCTACCAACACACCATGCTCCACCTGACTTTGGCTTGGAGACTG 1690  
Qy 1620 ATCTCTCTGTGTAAATTTCTGCCGGTGTGTGAAGGCTGGAAGGCTGGAAGGCTGCTGG 1679  
Db 1691 ATCTCTCTGTGTAAATTTCTGCCGGTGTGTGAAGGCTGGAAGGCTGGAAGGCTGCTGG 1750  
Qy 1680 GGTCTCTCTGGACCCGCTGTTGCTTCTGCCCTCCCTCTGGGAAAGGCTTATATGACGG 1739  
Db 1751 GGTCTCTCTGGACCCGCTGTTGCTTCTGCCCTCCCTCTGGGAAAGGCTTATATGACGG 1810  
Qy 1740 CCGCTCAGGCCCCAGAACTGCTGTGAACACACTTTTCCAGCCAGAGTTCCTCAAGCT 1799  
Db 1811 CCGCTCAGGCCCCAGAACTGCTGTGAACACACTTTTCCAGCCAGAGTTCCTCAAGCT 1870  
Qy 1800 GGAACGCTAGCTGCTCTTCTTAAAGATGSCCTTCCCTCCAGCCGACCGCCACGCTCA 1859  
Db 1871 GGAACGCTAGCTGCTCTTCTTAAAGATGSCCTTCCCTCCAGCCGACCGCCACGCTCA 1930  
Qy 1860 GTTGCCAGGATGGGCCCCACCTGCTCACTGTGGAATACAGACAGTGAATCTGTCT 1919  
Db 1931 GTTGCCAGGATGGGCCCCACCTGCTCACTGTGGAATACAGACAGTGAATCTGTCT 1990  
Qy 1920 GCCTGAACGAGTCAATTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGTCCCC 1979  
Db 1991 GCCTGAACGAGTCAATTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATAAGTCCCC 2050  
Qy 1980 TGACAGTGAATGTGTGGTGGGGGAGCCCTTGCCTCAAAAATTCACAGAGAAATGCC 2039  
Db 2051 TGACAGTGAATGTGTGGTGGGGGAGCCCTTGCCTCAAAAATTCACAGAGAAATGCC 2110  
Qy 2040 TCTCAGCCTCATGTGTGGTTCCTCTCTCTAGCTCCCGAGGATGTTGGGGACCCA 2099  
Db 2111 TCTCAGCCTCATGTGTGGTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2170  
Qy 2100 GCTTGTCTCGGACGTAAAGACAGTGAACAGATGTGAATTTGGGACCTGTGTGGTG 2159  
Db 2171 GCTTGTCTCGGACGTAAAGACAGTGAACAGATGTGAATTTGGGACCTGTGTGGTG 2230  
Qy 2160 GCCTTGAGCTGCTTCTGTGTGTGGAGACTGACTCCCATTTCTAAAGGAATGCC 2219  
Db 2231 GCCTTGAGCTGCTTCTGTGTGTGGAGACTGACTCCCATTTCTAAAGGAATGCC 2290  
Qy 2220 CGGGAGGACATTTGGAGGAAGATGSCCTGAGTGTGCACTTTGGCTTCTGCTACCTGCTCC 2279  
Db 2291 CGGGAGGACATTTGGAGGAAGATGSCCTGAGTGTGCACTTTGGCTTCTGCTACCTGCTCC 2350  
Qy 2280 TGAAGCCCGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2339  
Db 2351 TGAAGCCCGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2410  
Qy 2340 ACTGGAGCTCTAGAGAGCTGGGCTGTGTGTCTTTTGGCCCTTTTGTCTTACCTAAAT 2399  
Db 2411 ACTGGAGCTCTAGAGAGCTGGGCTGTGTGTCTTTTGGCCCTTTTGTCTTACCTAAAT 2470



QY 2400 GAAGAAACCATGCTGGAGGGCGGTGAACACAGAACCCCTCAGACAAAGATGACAGAC 2459  
 Db 2471 GAAGAAACCATGCTGGAGGGCGGTGAACACAGAACCCCTCAGACAAAGATGACAGAC 2530  
 QY 2460 TGGAGGACACATCTAGCTGCGCATTCGCAACCTCACTGGGCTCCCGACACTCTGTGTGAG 2519  
 Db 2531 TGGAGGACACATCTAGCTGCGCATTCGCAACCTCACTGGGCTCCCGACACTCTGTGTGAG 2590  
 QY 2520 AATTTAAACCCCTGCTTGTGAG 2544  
 Db 2591 AATTTAAACCCCTGCTTGTGAG 2615

RESULT 6

AAI60518/C  
 ID AAI60518 standard; cDNA; 2380 BP.  
 XX  
 AC AAI60518;  
 XX  
 DT 22-OCT-2001 (first entry)  
 XX  
 DE Human polynucleotide SEQ ID NO 4507.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; SS.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 XX 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0682191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Auandi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao Q, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR P-PSDB; AAM41362.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Claim 1; SEQ ID NO 4507; 1007app; English.  
 XX  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SQ Sequence 2380 BP; 510 A; 664 C; 670 G; 536 T; 0 other;  
 Query Match 92.9%; Score 2374.2; DB 22; Length 2380;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 2376; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 QY 176 CTTCTGTGCAACCTGGGAAACAGTGGATGCCACCCCATTTGGTGTGTGCCATCAACAG 235  
 Db 2379 CTTCTGTGTGCAACCTGGGAAACAGTGGATGCCACCCCATTTGGTGTGTGCCATCAACAG 2320  
 QY 236 CAAGATCCTAGTGGCCCCCAGGACCATCCATGCCAAGTTCACCGACTGGAGGGCTACCT 295  
 Db 2319 CAAGATCCTAGTGGCCCCCAGGACCATCCATGCCAAGTTCACCGACTGGAGGGCTACCT 2260  
 QY 296 CATGAACCGCTGGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGA 355  
 Db 2259 CATGAACCGCTGGTGGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGA 2200  
 QY 356 GAGCATGAAGTACCCCTTTAGGACGGGATCGGCTGGAGTGGTGGACAGTCCCGAGT 415  
 Db 2199 GAGCATGAAGTACCCCTTTAGGACGGGATCGGCTGGAGTGGTGGACAGTCCCGAGT 2140  
 QY 416 GTCACGCACTCGCATGGCTGTGTGTGGACACAGTAAATCGGGGTCGCTTACGGCTCTCTA 475  
 Db 2139 GTCACGCACTCGCATGGCTGTGTGTGGACACAGTAAATCGGGGTCGCTTACGGCTCTCTA 2080  
 QY 476 CGAGATGGTGACAGTGACAGCACTTCTGGTGCACATGTGGAGAGCCCTGATCCACCC 535  
 Db 2079 CGAGATGGTGACAGTGACAGCACTTCTGGTGCACATGTGGAGAGCCCTGATCCACCC 2020  
 QY 536 AGTGGGTTGGTACAGCGTGTGGGCCACGGCATCAAGATGTCAGAGAGGCGAAGTGACAT 595  
 Db 2019 AGTGGGTTGGTACAGCGTGTGGGCCACGGCATCAAGATGTCAGAGAGGCGAAGTGACAT 1960  
 QY 596 GGCCCATCACCCCACTTCCGGAAAGATCTACTGTGATGCCGTTCTTACCTCTTCAAGAA 655  
 Db 1959 GGCCCATCACCCCACTTCCGGAAAGATCTACTGTGATGCCGTTCTTACCTCTTCAAGAA 1900  
 QY 656 GGTACGACAGTCTACACAGAGGGGTTGGTTTGGAGAGGGATGAAGCTGGAGGCCAT 715  
 Db 1899 GGTACGACAGTCTACACAGAGGGGTTGGTTTGGAGAGGGATGAAGCTGGAGGCCAT 1840  
 QY 716 TGACCCCTGAATCTGGGCAACATCTGCGTGGCAACTGTCTGTAAAGTTCTCTGGATGG 775  
 Db 1839 TGACCCCTGAATCTGGGCAACATCTGCGTGGCAACTGTCTGTAAAGTTCTCTGGATGG 1780  
 QY 776 ATACCTGATGATCTGTGTGGAGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGCTA 835  
 Db 1779 ATACCTGATGATCTGTGTGGAGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGCTA 1720  
 QY 836 CCATGCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGAGAGAAATGACATTTAGCT 895  
 Db 1719 CCATGCTCTTCCCAAGCCATCTTCCCGGCCACCTTCTGTGAGAGAAATGACATTTAGCT 1660  
 QY 896 CACACCGCCAAAAGTTATAGGACAGACTTTCACTGGGAGAACTACTTTGGAGAGAC 955  
 Db 1659 CACACCGCCAAAAGTTATAGGACAGACTTTCACTGGGAGAACTACTTTGGAGAGAC 1600  
 QY 956 CAAAGTCGAAAGCCGCTCCATCGAGACTCTTTAAATGATTTGCCAACCACTGGCTTCAA 1015  
 Db 1599 CAAAGTCGAAAGCCGCTCCATCGAGACTCTTTAAATGATTTGCCAACCACTGGCTTCAA 1540  
 QY 1016 GGTGGGCAATGAAGCTGGAGGGCCGTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCAC 1075  
 Db 1539 GGTGGGCAATGAAGCTGGAGGGCCGTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCAC 1480  
 QY 1076 GGTGAACAGTGTGTGTCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTA 1135

Db 1479 GGTGAACAGTGTGTGATCGCTTCCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTA 1420  
 QY 1136 CGACAGTGGTGGAGTCCGAGTCCCGAGACATACCTCCCGTGGCTGGTGTGAGTCCAC 1195  
 Db 1419 CGACAGTGGTGGAGTCCGAGTCCCGAGACATACCTCCCGTGGCTGGTGTGAGTCCAC 1360  
 QY 1196 CGGTACAGTCCAGCTCCTGTGGCGGAGAACCGGACACACCGCTGAAGGCCAAGA 1255  
 Db 1359 CGGTACAGTCCAGCTCCTGTGGCGGAGAACCGGACACACCGCTGAAGGCCAAGA 1300  
 QY 1256 GGCACAAAGAAAGAAAGAAAGTGGGAGAAAGAAAGAAAGTCCCGCCACTAA 1315  
 Db 1299 GGCACAAAGAAAGAAAGAAAGTGGGAGAAAGAAAGTCCCGCCACTAA 1240  
 QY 1316 GACGGACCCCTCAGACAGGGGTCCAGAGCCCTGTGGAGAGACACCTCAGGGTGC 1375  
 Db 1239 GACGCGACCCCTCAGACAGGGGTCCAGAGCCCTGTGGAGAGACACCTCAGGGTGC 1180  
 QY 1376 CAGGAAGATCTCGTGGAGCTGTCTTCTGGGAGATCATTTGCTGTGGTGTGAAGGAAGA 1435  
 Db 1179 CAGGAAGATCTCGTGGAGCTGTCTTCTGGGAGATCATTTGCTGTGGTGTGAAGGAAGA 1120  
 QY 1436 GCATCTAGACGTGGCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCTCTCCGTCGA 1495  
 Db 1119 GCATCTAGACGTGGCTCGCCGACAAAGCTTCAAGTCCAGAGCTGCTCTCCGTCGA 1060  
 QY 1496 GAAATCATCAGCAGGAAACAGAGCTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGG 1555  
 Db 1059 GAAATCATCAGCAGGAAACAGAGCTGAGCTTCTGCTCCAGCTGGCTTCTAGCTGG 1000  
 QY 1556 AAGCAGCAGCAGCTTCTTACACACACACACCTGCTCCACCTGACTTTGGCTTGGAG 1615  
 Db 999 AAGCAGCAGCAGCTTCTTACACACACACACCTGCTCCACCTGACTTTGGCTTGGAG 940  
 QY 1616 ACTGATCTCTCTGTGTAATTTCTGCGGCTGTGTGAAGCTGGAAGCTGTGAGGACCTG 1675  
 Db 939 ACTGATCTCTCTGTGTAATTTCTGCGGCTGTGTGAAGCTGGAAGCTGTGAGGACCTG 880  
 QY 1676 CTGGGCTCTCTGGGACCGCTGTGTCTTCTGCTCCCTCCCTGGAAGCTGTATATGA 1735  
 Db 879 CTGGGCTCTCTGGGACCGCTGTGTCTTCTGCTCCCTCCCTGGAAGCTGTATATGA 820  
 QY 1736 CGGCGCTCTGAGCAGCAGCAGTCTGTGTGAACCACTTTTCCAGCAGAGTTCOCAC 1795  
 Db 819 CGGCGCTCTGAGCAGCAGCAGTCTGTGTGAACCACTTTTCCAGCAGAGTTCOCAC 760  
 QY 1796 AGCTGGAACGTAGTGTGCTCTTCTTCTTAAAGTGGCTCCCGGACCGCCAGCGCC 1855  
 Db 759 AGCTGGAACGTAGTGTGCTCTTCTTCTTAAAGTGGCTCCCGGACCGCCAGCGCC 700  
 QY 1856 CTCAGTTGCCAGGATGGGGCCACCACTGTGCACACTGTGGAATACAAGACAGTGAATCT 1915  
 Db 699 CTCAGTTGCCAGGATGGGGCCACCACTGTGCACACTGTGGAATACAAGACAGTGAATCT 640  
 QY 1916 GTCTGCTGACAGATCTGTAAATTAAGTCTAGAGCAGCTCTCTGAGGAGTAAGGT 1975  
 Db 639 GTCTGCTGACAGATCTGTAAATTAAGTCTAGAGCAGCTCTCTGAGGAGTAAGGT 580  
 QY 1976 CCCTGACAGTGTGTGTGGGGGAGCTGTGCTTCAAAATTCACCAAGCAGAA 2035  
 Db 579 CCCTGACAGTGTGTGTGGGGGAGCTGTGCTTCAAAATTCACCAAGCAGAA 520  
 QY 2036 TGCCTCTCAGCTCAGTGTGGTCTCTGCTCTCTCTAGTCCCGGAGTGTGGGA 2095  
 Db 519 TGCCTCTCAGCTCAGTGTGGTCTCTGCTCTCTCTAGTCCCGGAGTGTGGGA 460  
 QY 2096 CCCAGTGTCTCGGACGTAGAGCAGTGTGAGGAGTGTGGATTTGGGACCTGTGT 2155  
 Db 459 CCCAGTGTCTCGGACGTAGAGCAGTGTGAGGAGTGTGGATTTGGGACCTGTGT 400  
 QY 2156 GGTGGCTTGAAGTGTCTGTGTGTGTGGAGTGAAGTCTCCCTTCTTAAAGGAATG 2215

Db 399 GGTGGCTTGAAGTGTCTGTGTGTGTGGAGTGAAGTCCCTTCTTAAAGGAATG 340  
 QY 2216 CCCCCTGGGAGGACATTTGGGAGGAAGATGGCTGTGAGTGTGCACTTGGCTCTGTACCTG 2275  
 Db 339 CCCCCTGGGAGGACATTTGGGAGGAAGATGGCTGTGAGTGTGCACTTGGCTCTGTACCTG 280  
 QY 2276 CTTCTTGAAGCCCGCTTAAATTAATTCATCAAGATTCCTTTGTAGTTAAAGGGTCCAGT 2335  
 Db 279 CTTCTTGAAGCCCGCTTAAATTAATTCATCAAGATTCCTTTGTAGTTAAAGGGTCCAGT 220  
 QY 2336 TCTGACTGGAGCTCTAGAGAGCTGGGCTTGTATGTTCTTTTGGCTTTTGTCTACT 2395  
 Db 219 TCTGATTTGGAGCTCTAGAGAGCTGGGCTTGTATGTTCTTTTGGCTTTTGTCTACT 160  
 QY 2396 AATGAGAAACCATGCTGGAGGGCGGTGAACACAGACCTCAAGACAGGATGACA 2455  
 Db 159 AATGAGAAACCATGCTGGAGGGCGGTGAACACAGACCTCAAGACAGGATGACA 100  
 QY 2456 GAGCTGGAGGACACATCTAGCTGCTTGAACCTCACTGGGCTCCCGAGACTCTGTGTG 2515  
 Db 99 GAGCTGGAGGACACATCTAGCTGCTTGAACCTCACTGGGCTCCCGAGACTCTGTGTG 40  
 QY 2516 TGAGAAATTAACCCCTGCTTGTGTTGAGAAAAA 2554  
 Db 39 TGAGAAATTAACCCCTGCTTGTGTTGAAAAA 1

RESULT 7

ABA93718  
 ID ABA93718 standard; cDNA; 3071 BP.  
 XX ABA93718;  
 AC AC  
 DT DT  
 XX 30-APR-2002 (first entry)  
 DE Human transmembrane protein cDNA clone amy2\_1114.  
 XX Human; foetal brain; foetal kidney; melanoma; testis; amygdala;  
 KW gene therapy; chromosome 22q13.1-13.33; ss.  
 XX Homo sapiens.  
 OS  
 XX WO200198454-A2.  
 PN  
 XX 27-DEC-2001.  
 PD  
 XX 25-APR-2001; 2001WO-IB02050.  
 PF  
 XX 25-APR-2000; 2000US-199380P.  
 PR  
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 PA  
 XX Wiemann S;  
 PI  
 XX WPI; 2002-055860/07.  
 DR P-PSDB; AB805680.  
 DR  
 XX Human cDNA sequences and clones derived from human fetal brain, fetal  
 PT kidney, melanoma, testis and amygdala cDNA libraries, useful in genetic  
 PT screening and therapy -  
 XX  
 PS Claim 1; Page 128-129; 611pp; English.

CC The present invention describes assemblies and computer readable media  
 CC comprising novel human cDNA sequences and clones derived from human  
 CC foetal brain, foetal kidney, melanoma, testis and amygdala cDNA  
 CC libraries. ABA93702 to ABA93766 represent human cDNA sequences from the  
 CC present invention which encode the proteins given in ABB05662 to  
 CC ABB05729. The human cDNA sequences and clones can be used in gene  
 CC therapy. The clones may be used in a variety of applications, for  
 CC example they may be used in profiling assays, for providing large arrays  
 CC of human genetic material for implementing large-scale screening  
 CC strategies and for treating diseases via gene therapy procedures.





QY 1621 TCCTCTCTGTAAATCTGCCGGTGTGTAAGGCTGACGCTGGAGGACCTGTCTGG 1680  
Db 2750 TCCTCTCTGTAAATCTGCCGGTGTGTAAGGCTGACGCTGGAGGACCTGTCTGG 2809  
QY 1681 GTCTCTGGGACCGCGCTGTCTGCTTCCCTCCCTGCGAAAGTCTATATGACGGG 1740  
Db 2810 GTCTCTGGGACCGCGCTGTCTGCTTCCCTCCCTGCGAAAGTCTATATGACGGG 2869  
QY 1741 CGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTG 1800  
Db 2870 CGCTGAGGCCCCAGAACTGCTGTGAACCACTTTTCCAGCCAGAGTTCCTCAAGCTG 2929  
QY 1801 GAACGCTAGCTGCTGCTTCTCTTAAGATGGCTTCCCGCGACCCCGCCAGGCGCTCAG 1860  
Db 2930 GAACGCTAGCTGCTGCTTCTCTTAAGATGGCTTCCCGCGACCCCGCCAGGCGCTCAG 2989  
QY 1861 TTGCNAGGATGGGCGCACCACTGTACACTGTGGAATACAGACAGTGAACCTGTCTG 1920  
Db 2990 TTGCNAGGATGGGCGCACCACTGTACACTGTGGAATACAGACAGTGAACCTGTCTG 3049  
QY 1921 CCTGAACGAGTCATGTAA 1939  
Db 3050 CCTAAGAAAAA 3068

RESULT 8  
AAZ94124  
ID AAZ94124 standard; cDNA; 3590 BP.  
XX AAZ94124;  
AC AAZ94124;  
XX 19-JUN-2000 (first entry)  
XX Haematopoietic stem cell specific nucleic acid.  
XX Haematopoietic stem cell; immune system disorder;  
KW leukaemia; antileukemic; immunomodulator; therapy; mouse; ss.  
XX Mus musculus.  
XX  
FH Key Location/Qualifiers  
FT CDS 260..2155  
FT /tag= a  
FT /transl\_except= (pos:1748..1750, aa:Val)  
FT /transl\_except= (pos:1826..1828, aa:Arg)  
FT /transl\_except= (pos:1901..1903, aa:Asn)  
XX  
XX WO200011168-A2.  
XX  
XX 02-MAR-2000.  
XX 20-AUG-1999; 99WO-US19052.  
XX 21-AUG-1998; 98US-0138132.  
XX (UYPR-) UNIV PRINCETON.  
XX  
XX Lemischka I, Moore K;  
XX WPI; 2000-237650/20.  
XX P-PSDB; AAY79183.  
XX  
XX Hematopoietic stem cell signaling proteins modulating replication and  
XX differentiation for treating immune system disorders and leukaemia -  
XX Claim 10; Page 228-229; 256pp; English.  
XX  
XX The present sequence is that of a nucleic acid isolated from  
XX mouse primitive stem cells by a method of the invention. It is an  
XX example of claimed isolated nucleic acids (see AAZ94077-294131) that  
XX are specifically expressed in haematopoietic stem cells (HSCs) and  
XX which encode HSC-specific proteins. The HSCs are especially

CC primitive HSCs (PHSCs) such as umbilical cord cells, bone marrow  
CC cells and foetal liver cells. The encoded proteins (see AAY79176-93)  
CC are growth factors, transcription factors, splicing factors, capping  
CC factors, transport proteins, translation factors or replication  
CC factors that modulate HSC activity, especially differentiation or  
CC replication. The invention provides a claimed method for  
CC identifying PHSC-specific nucleic acids involving: creating a PHSC  
CC cDNA library and a non-PHSC immune cell library; and subtracting  
CC the 2 libraries. Also claimed are methods for generating a stem  
CC cell/progenitor cell from PHSCs; for identifying the presence of a  
CC PHSC in a sample; for identifying the presence in a sample of a  
CC compound that modulates HSC activity; for using such a compound to  
CC treat an immune system condition, especially leukaemia; for  
CC introducing exogenous nucleic acid into a HSC; and for ex vivo  
CC expansion of HSCs. Also claimed is a PHSC specifically expressing 1  
CC of the claimed nucleic acids, such as the present sequence.  
XX  
SQ Sequence 3590 BP; 1101 A; 702 C; 797 G; 990 T; 0 other;  
Query Match 16.2%; Score 413.8; DB 21; Length 3590;  
Best Local Similarity 60.2%; Pred. No. 6.1e-104;  
Matches 735; Conservative 0; Mismatches 462; Indels 24; Gaps 2;  
QY 1 AGCACGCTCCACTCTATGACAGTGGAGGATGTGATGAAGGGATGAAGTGGAGTGC 60  
Db 750 AGCATGCACCTATGGGACCTCTGCGGTGATATCTCGAAATGTAAGGATAGAAGTTC 809  
QY 61 TCAACAGTATGCTGTGCTCCCGCGGGTGTACTGGATGCCCTCTGTCAATCCAGACAG 120  
Db 810 CCAATACAGACTGCGAGTCTACCTGCAAGTCTCTGGATTCTGGAATATAAATAG 869  
QY 121 CAGGGTATCGGTGCTTTCGGTATGAAGGCTTTTGAATAACGCCAGCCAGTCTTCT 180  
Db 870 CAGGTATATAATGCCCTTTTGAGATATGAAGGATTTGAAATGATCTCTCTGGACTTCT 929  
QY 181 GTTGCAACCTGGGAACAGTGGATGTCACCCCATTTGGTGGTGGTCCCATCAACAGCA 240  
Db 930 GTTGCAATATATGCGGTCTGATATTCATCCAGTGTGGTGGTGGTGGTGGTGGTGG 989  
QY 241 TCCTAGTGGCCCGACCGGACCATCCATGCCAAGTTCACCGACTGGAAGGGCTACCTCA 300  
Db 990 CTCTGCTTCTCTAGAACTGTTCAACATAATATACAACTGGAAGCTTTTCTAGTAA 1049  
QY 301 AACGGCTGTGGCTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGCA 360  
Db 1050 AAGACTTACTGTGTCACAAACACCTTCTCTCTGATTTCTCAAAAGGTTTCTGAG 1109  
QY 361 TGAAGTACCCCTTTAGCGAGGATGCGGCTGCAAGTGTGACAAAGTCCCGAGTGTAC 420  
Db 1110 TGCAATATCTTTCAAACTTGCATGAGATGAGTAGTAGTGTACAGAGGCAATTTAT 1169  
QY 421 GCACCTCGCATGGCTGTGGTGGACACAGTAATCGGGGTGCGCTACGGCTCTCTAGCAG 480  
Db 1170 GAACAAGAGTAGCAGTGGTGGAAAGTAAATGGAGGACGACTACGGCTGGTATGAAG 1229  
QY 481 ATGGTGACAGTG---ACGAGGACTTCTGGTGGACATGTGGAGCCCGCTGATCCACCC 537  
Db 1230 AGAGTGAGATGGAAACAGACGACTTCTGGTGGACATGACAGCCCGCTTATTCACCA 1289  
QY 538 TGGGTTGGTCAACGCTGTGGGCGGCGCATCAAGATGTCAAGAGGCGGAAGTGAATGG 597  
Db 1290 TTGATGGTCAAGAGCATAGGCCATCGATTCAAGAGATCAGATATTACGAAGAAACAG 1349  
QY 598 CCATCAACCCACCTTCCCGAAGATCTACTGTATGCCGTTCTCTTACCTCTTCAAGAG 657  
Db 1350 ACGGAC-----ATTTCGATACACCTCCACACTTATTTCCTAAGG 1388  
QY 658 TACGACGCTTACACAGAGGCGGTTGTTTGGAGGAAGGATGAAGCTGGAGGCCATTG 717  
Db 1389 TAAAGAGTAGACAGAGTGGAGATGTTCAAGAGGAATGAATTTGGAAGCTATAG 1448  
QY 718 ACCCCCTGAACTGGGCAACATCTGCGTGGCAACTCTCTGTAAAGTTCTCTCGATGGAT 777

1449 ACCATTAAATCTTTTACAAATATGTGTTGCCACCAATTAGAAAGGTCTGGCTGATGGAT 1508  
 778 ACCTGATGATCTGTGTGACGGGGCCCTCCAGAGATGGCTTGGACTGCTTCTGTACC 837  
 1509 TCCTGATGATGGGATGATGGCTCAGAGCAGAGATGGATCTGACTGGTCTGTATC 1568  
 838 ATGCTCTTCCAGGCCATCTTCCGGGCCACCTTCTGTGAGAGATGACATGGAGTCA 897  
 1569 ATGCAACCTCTCTTCCATTTTCCCTGTGGGTTTCTGTGAAATTAACATGATAGAACTGA 1628  
 898 CACGCCAAAGGTTATGAGCAGCAGATTTCACTGGGAGACTTCTGGAGAGACCA 957  
 1629 CTCACCCAGAGGTTACAAAACTTCTTTTAAATGGTTTGAATCTCAGGGAACCG 1688  
 958 AGTCGAAAGCCGCTCCATCGAGATCTTTTAACTGATTTGCCAAACCATGGCTTCAAGG 1017  
 1689 GCTCATTTGACGACCACTAAACTATTTAATAAGGATTTTCCAAACCAACGATTCGGT 1748  
 1018 TGGGATGAGCTGGAGCCGCTGGACCTGATGGAGCCCGCTCATCTGTGTGCCACGG 1077  
 1749 TAGGAATGAAATTAAGAGCTGTAGATCTCATGGAGCCACCGTTAATATGTGTAGCCACAG 1808  
 1078 TGAACAGAGTGTGATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACG 1137  
 1809 TTACTCGAATTAATCACCATCTCTTGAGATACATTTTGATGGTGGAGAGAGATG 1868  
 1138 ACCAGTGGGTGGAGTGGAGTCCCGAGACATCTACCCCGTGGCTGTGTGAGCTACCG 1197  
 1869 ACCAGTGGGTGAGTGTGAGTCCCGTACCTCTATCTCTGAGGGTGGTGTGAGTAACTG 1928  
 1198 GCTACCGCTCCAGCTCCCTG 1218  
 1929 GATATCAACTACGCTCCAG 1949

RESULT 9  
 ID ABX08841  
 XX ABX08841 standard, cDNA; 2564 BP.  
 AC ABX08841;  
 XX  
 XX  
 DT 21-JAN-2003 (first entry)  
 XX  
 DE Angiogenesis-associated human polynucleotide sequence #103.  
 XX  
 KW Human; angiogenesis-associated transcript; angiogenesis;  
 KW angiogenesis-associated disease; cancer; cytostatic; gene therapy;  
 KW Gene; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200279492-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 XX 14-FEB-2002; 2002WO-US04915.  
 XX  
 XX 14-FEB-2001; 2001US-0784356.  
 PR 22-FEB-2001; 2001US-0791390.  
 PR 19-APR-2001; 2001US-285475P.  
 PR 03-AUG-2001; 2001US-31025P.  
 PR 13-NOV-2001; 2001US-350666P.  
 PR 29-NOV-2001; 2001US-334244P.  
 XX  
 XX (EOSB-) EOS BIOTECHNOLOGY INC.  
 XX  
 XX Murray R, Glynn R, Watson SR, Aziz N;  
 PI  
 XX WPI; 2003-040681/03.  
 DR P-PSDB; ABU03557.  
 DR  
 XX Detecting angiogenesis-associated transcript in a cell for diagnosing  
 PT and treating cancer by contacting a sample with a polynucleotide that

PT exhibits changes in expression level as a function of time in tissue  
 PT undergoing angiogenesis  
 XX  
 XX Example 2; Page 274-275; 291pp; English.  
 XX  
 CC The present invention relates to methods and compositions for  
 CC detecting an angiogenesis-associated transcript in a cell in  
 CC a patient. The method involves contacting a biological sample from  
 CC the patient with a polynucleotide that selectively hybridizes to a  
 CC sequence at least 80% identical to any of the angiogenesis-associated  
 CC human polynucleotide sequences given in the specification. These  
 CC angiogenesis-associated polynucleotide sequences comprise genes that  
 CC exhibit changes in expression levels as a function of time in tissue  
 CC undergoing angiogenesis. The method and the polynucleotide sequences  
 CC of the invention are useful for diagnosing and treating angiogenesis  
 CC and angiogenesis-associated diseases e.g. cancer. The polynucleotide  
 CC sequences are also useful in the gene therapy of such disorders. The  
 CC angiogenesis-associated proteins encoded by the polynucleotide  
 CC sequences are useful as a vaccine for therapeutic and prophylactic  
 CC immunisation. ABX08739-ABX08853 represent angiogenesis-associated  
 CC polynucleotide sequences.  
 XX

SQ Sequence 2564 BP; 832 A; 482 C; 560 G; 690 T; 0 other;

Query Match 15.9%; Score 405.2; DB 25; Length 2564;

Best Local Similarity 60.2%; Pred. No. 1.3e-101;

Matches 721; Conservative 0; Mismatches 453; Indels 24; Gaps 2;

QY 24 TGGAGGATGTCATGAAAGGATGAAGGTGGAGGTGCTCAACAGTGATGCTGTGTCGCC 83  
 DB 181 TGGGGTGATATCTCAGAAAAATGTGAGAGTAGAAGTTCCCAATACAGACTGCGAGCCTACCT 240  
 QY 84 AGCCGGTGTACTGATCGCCTCTGTCATCCAGACAGAGGATATCGGGTGTGCTCGG 143  
 DB 241 ACCAAGTCTTCTGGAATGCTGGAATGTAAATTTAGAGGTTCAATGCCCTTTTAA 300  
 QY 144 TATGAAGGCTTTGAAATGACCCAGCCATGACTTCTTGGTGCAACTGGGAAACAGTGAT 203  
 DB 301 TATGAAGGATTTGAAATGACTCTGCTCGGACTTCTGGTGCAATATATGTGTTCTGAT 360  
 QY 204 GTCCACCCATTTGGTGTGTCCTCATCAACCAAGATCTTAGTCCCGCCCGACCCATC 263  
 DB 361 ATCCATCCAGTTGGTGTGTCAGCCAGCGAAACCTTTGTTCTCTTAGAATATTT 420  
 QY 264 CATGCCAAGTTCACCGACTGGAAGGCTACCTCATGAAACGGCTGGTGGCTTCCAGGACG 323  
 DB 421 CAGCATAATATACAACTGGAAAGCTTTCTAGTGAAGAGCTTACTTGTGCCAAACA 480  
 QY 324 CTTCCCGTGATTTCCATCAAGATGGTGGAGAGCATGAAGTACCCCTTTAGGACGGC 383  
 DB 481 CTGCTCTCTGATTTCTCCAAAAGGTTTCAGAGAGTATGCAATATCTTCAACCTTGC 540  
 QY 384 ATGCGGCTGGAAGTGGTGGACCAAGTCCAGGTGTCAACGCTCTCGATGGCTGTGGAC 443  
 DB 541 ATGAGAGTAGAGTGGTTCAGAGAGGATTTGTGTGCAACACGAGTAGCAGTGGTGA 600  
 QY 444 ACAGTAATCGGGGTGCGCTACGGCTCTCTACAGGA---TGGTGACAGTGACACGAC 500  
 DB 601 AGTGTAAATTCGAGGAAGATTAAGACTAGTGTATGAAGAAAGCCAAAGATAGAACATGAC 660  
 QY 501 TTCTGGTGCCACATGTGGAGCCCGCTGATCCACCCAGTGGGTGGTTCAGACGCTGTGGC 560  
 DB 661 TTCTGGTGCCATATGACAGCCCATTAATACATCATATTTGGTGGTCTCGAAGCATAGT 720  
 QY 561 CACGGCATCAAGATGTTCAGAGAGGGGAAAGTGAATGAGGCCCATCACCCACCTTCCGAG 620  
 DB 721 CATCGATTCAAAAGATCTGATATTAACAAGAA-----ACAGGAT 759  
 QY 621 ATCTACTGTGATGCGCTTCTTACCTTCAAGAGGTACGAGCTGTACACAGAGGC 680  
 DB 760 GGACATTTTGTATACACCAACATTTATTTGTAGGTAAAGAGTAGACAGAGTGGG 819  
 QY 681 GGTGTTTGGAGGAGGATGAAGCTGGAGGCCAATTTGACCCCTGAATCTGGGCAACATC 740



XX 13-DEC-2001.  
 XX 30-MAY-2001; 2001WO-US10838.  
 XX 05-JUN-2000; 2000US-209473P.  
 XX 05-JUN-2000; 2000US-209531P.  
 XX 18-SEP-2000; 2000US-233133P.  
 XX 18-SEP-2000; 2000US-233617P.  
 XX 20-SEP-2000; 2000US-234009P.  
 XX 20-SEP-2000; 2000US-234034P.  
 XX 20-SEP-2000; 2000US-234052P.  
 XX 20-SEP-2000; 2000US-234059P.  
 XX 22-SEP-2000; 2000US-234509P.  
 XX 25-SEP-2000; 2000US-234923P.  
 XX 25-SEP-2000; 2000US-234924P.  
 XX 25-SEP-2000; 2000US-235077P.  
 XX 25-SEP-2000; 2000US-235082P.  
 XX 25-SEP-2000; 2000US-235134P.  
 XX 25-SEP-2000; 2000US-235280P.  
 XX 26-SEP-2000; 2000US-235637P.  
 XX 26-SEP-2000; 2000US-235638P.  
 XX 27-SEP-2000; 2000US-235711P.  
 XX 27-SEP-2000; 2000US-235720P.  
 XX 27-SEP-2000; 2000US-235840P.  
 XX 27-SEP-2000; 2000US-235863P.  
 XX 28-SEP-2000; 2000US-236032P.  
 XX 28-SEP-2000; 2000US-236033P.  
 XX 28-SEP-2000; 2000US-236034P.  
 XX 28-SEP-2000; 2000US-236109P.  
 XX 28-SEP-2000; 2000US-236111P.  
 XX 29-SEP-2000; 2000US-236842P.  
 XX 29-SEP-2000; 2000US-236891P.  
 XX 02-OCT-2000; 2000US-237172P.  
 XX 02-OCT-2000; 2000US-237173P.  
 XX 02-OCT-2000; 2000US-237278P.  
 XX 02-OCT-2000; 2000US-237294P.  
 XX 02-OCT-2000; 2000US-237295P.  
 XX 02-OCT-2000; 2000US-237315P.  
 XX 03-OCT-2000; 2000US-237425P.  
 XX 03-OCT-2000; 2000US-237598P.  
 XX 03-OCT-2000; 2000US-237604P.  
 XX 03-OCT-2000; 2000US-237606P.  
 XX 03-OCT-2000; 2000US-237608P.  
 XX 01-NOV-2000; 2000US-244867P.  
 XX 01-NOV-2000; 2000US-245084P.  
 XX (AVAL-) AVALON PHARM.  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 XX Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 XX Screening for anti-neoplastic agent involves exposing cells to a  
 XX chemical agent to be tested for anti-neoplastic activity, and  
 XX determining a change in expression of a gene of a signature gene set -  
 XX Claim 1; SEQ ID 7266; 44pp; English.  
 XX The present invention describes a method (M1) for screening for an  
 XX anti-neoplastic agent. The method involves exposing cells to a chemical  
 XX agent to be tested for anti-neoplastic activity, determining a change in  
 XX expression of at least one gene (I) of a signature gene set, where (I)  
 XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 XX to ABL70110), or is at least 95% identical to (S), where a change in  
 XX expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 XX activity and can be used in gene therapy. M1 can be used for screening  
 XX an anti-neoplastic agent, and can be used for producing a product which  
 XX is the data collected with respect to the anti-neoplastic agent as a  
 XX result of M1, and the data is sufficient to convey the chemical  
 XX structure and/or properties of the agent. M1 can be used in the

CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC esophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilm's tumour.  
 XX  
 SQ Sequence 418 BP; 103 A; 112 C; 124 G; 76 T; 3 other;  
 Query Match 11.0%; Score 281.4; DB 24; Length 418;  
 Best Local Similarity 97.3%; Pred. No. 1.2e-67;  
 Matches 285; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 1001 AAACCATGGCTTCAAGCTGGGCGCATGAGCTGGAGGCGGTGACCTGATGAGCCCGGCT 1060  
 DB 1 AAACCATGGCTTCAAGCTGGGCGCATGAGCTGGAGGCGGTGACCTGATGAGCCCGGCT 60  
 QY 1061 CATCTGTGTGGCCACCGGTGAAACGAGTGTGTCATCGGCTCTCAGCATCCACTTTGACGG 1120  
 DB 61 CATCTGTGTGGCCACCGGTGAAACGAGTGTGTCATCGGCTCTCAGCATCCACTTTGACGG 120  
 QY 1121 CTGGACAGCGAGTACACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1180  
 DB 121 CTGGACAGCGAGTACACGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180  
 QY 1181 CTGGTGTGAGCTACCCGGCTACCGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCC 1240  
 DB 181 CTGGTGTGAGCTACCCGGCTACCGCTCCAGCTCCAGCTCCAGCTCCAGCTCCAGCTCC 240  
 QY 1241 GCTCAAGGCCCAAGAGGCCCAAGAGGCCCAAGAGGCCCAAGAGGCCCAAGAGGCCCA 1293  
 DB 241 GCTCAAGGCCCAAGAGGCCCAAGAGGCCCAAGAGGCCCAAGAGGCCCAAGAGGCCCA 293  
 RESULT 12  
 AAH87128/c  
 ID AAH87128 standard; DNA; 289 BP.  
 XX  
 AC AAH87128;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Human single nucleotide polymorphism containing DNA sequence #1985.  
 XX  
 KW Biallelic marker; polymorphism; human; disease; diagnosis; treatment;  
 KW phenotypic trait; gene therapy; forensic; paternity; mapping; cancer;  
 KW transgenic; single nucleotide polymorphism; SNP; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Variation replace(126,A)  
 FT FT /\*tag= a  
 FT FT /standard\_name= "single nucleotide polymorphism"  
 XX  
 PN WO9953095-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 30-MAR-1999; 99WO-US06893.  
 XX  
 PR 09-APR-1998; 98US-0057871.  
 XX  
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.  
 XX  
 PI Lander ES, Wang D, Hudson T;  
 XX  
 DR WPI; 1999-620443/53.  
 XX  
 PT Polymorphic human genomic sequences and related allele-specific probes  
 PT and primers, useful for genetic analysis, e.g. diagnosis and monitoring  
 PT of disease -  
 XX  
 PS Claim 1; Page 243; 330pp; English.

XX This invention describes novel human nucleic acid segments (I)  
 CC containing polymorphic sites. The polynucleotides of (I) are used for,  
 CC e.g. correlating disease polymorphisms (or disease susceptibility) or  
 CC other phenotypic traits (e.g. baldness, obesity, fertility, strength,  
 CC response to drugs etc.); diagnosing and monitoring e.g. cancer,  
 CC inflammation, heart or central nervous system diseases; detecting  
 CC susceptibility to microbial infection; treating or preventing such  
 CC diseases; forensic analysis; gene therapy; paternity testing; mapping  
 CC genomic loci associated with phenotypic traits (and subsequent cloning  
 CC of the genes responsible); and the production of transgenic organisms.  
 CC Antibodies raised against (I) are useful as diagnostic and therapeutic  
 CC tools and in drug screening. AAH85144 - AAH87644 represent the human  
 CC DNA sequences containing biallelic polymorphic sites described in the  
 CC invention.  
 XX  
 SQ Sequence 289 BP; 72 A; 77 C; 68 G; 72 T; 0 other;  
 Query Match 10.6%; Score 271.8; DB 20; Length 289;  
 Best Local Similarity 98.6%; Pred. No. 4.6e-65;  
 Matches 285; Conservative 0; Mismatches 2; Indels 2; Gaps 1;  
 QY 2180 TTGTGAGGACTGACTGCC--ATTTCCTAAAGGAATGCCCGGGGAGGACATTTGGGAG 2237  
 DB 289 TTGTGAGGACTGACTGCCATTTCCCTAAAGGAATGCCCGGGGAGGACATTTGGGAG 230  
 QY 2238 GAAGATGCGCTGAGTGTCATTTGGCTCTGCTACCTGCTCTGAAACCCCGCTAAAT 2297  
 DB 229 GAAGATGCGCTGAGTGTCATTTGGCTCTGCTACCTGCTCTGAAACCCCGCTAAAT 170  
 QY 2298 AATTATCCCAAGATTCCTTTGTAGTAAAGGTCAGTTCTGACTGGAGCTCTAGAGAG 2357  
 DB 169 AATTATCCCAAGATTCCTTTGTAGTAAAGGTCAGTTCTGACTGGAGCTCTAGAGAG 110  
 QY 2358 CTGGGCTTGATGTTCTTTGGCCCTTTGTTCTTACCTAAATGAAGAAACCATGCTTGA 2417  
 DB 109 CTGGGCTTGATGTTCTTTGGCCCTTTGTTCTTACCTAAATGAAGAAACCATGCTTGA 50  
 QY 2418 GGGGCGGTGACACAGACACCTCAAGACAGGATGACAGAGCTGGAGGA 2466  
 DB 49 GGGGCGGTGACACAGACACCTCAAGACAGGATGACAGAGCTGGAGGA 1  
 RESULT 13  
 ABL27733  
 ID ABL27733 standard; DNA; 3486 BP.  
 XX  
 AC ABL27733;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 34672.  
 DE  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 FI  
 DR WPI; 2001-656860/75.  
 XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Claim 1; SEQ ID NO 34672; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511); expressed DNA  
 CC sequences (ABJ01840-ABL16175) and the encoded proteins  
 CC (ABJ57737-ABJ72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 3486 BP; 939 A; 901 C; 869 G; 777 T; 0 other;  
 Query Match 10.4%; Score 256.4; DB 23; Length 3486;  
 Best Local Similarity 55.6%; Pred. No. 5e-63;  
 Matches 678; Conservative 0; Mismatches 501; Indels 41; Gaps 7;  
 QY 92 GACTGGATCGCTCTGTCTATCCACAGCAGGATCGGTGCTTGGTATGAAGG 151  
 DB 1198 GTTTTGGTGGCCACATCTCTGGAATCAAGGCTTAAGGCTTAATGAGCTACAGGG 1257  
 QY 152 CTTTGAATAGCGCCAGCCATGACTTCTGTGCAACTGGGAACAGTGGATGTCCACC 211  
 DB 1258 TTTGATACGGACTC---GCAGACTTCTGGTGAACCTCTGCAATGCCAGGTGCAATC 1314  
 QY 212 CATTGGCTGGTGTGCAATCAGCAAGATCCTAGTGGCCCCCAGCAGCAGCAGTCCCAA 271  
 DB 1315 GGTGGTGTGGTGGCCATCTGGGGCAAGCCATTAATCCGGCCCCCAGCAGTGGACAA 1374  
 QY 272 GTTCCACCACTGGAAGGCTACTCATGAACGGCTGTGGGCTCCAGGAGCGTTCCTCGT 331  
 DB 1375 GTACAGGACTGGAAGGACTTCTTGGTGGAGCTTTATCCGAGGCCCGACCTTCCTC 1434  
 QY 332 GGAATTCACATCAAGATGGTGGAGAGCATGAAGTACCCCTTTAGGAGGAGCATCGGCT 391  
 DB 1435 CAACTTTTACACAAAATCAACGACAGCTCCAGTCGGCTTCGGCTTGGCGTGAATCT 1494  
 QY 392 GGAAGTGGTGGCAAGTCCAGGTCTCAGCATCGCATGGCTGTGGTGGACACAGTAAT 451  
 DB 1495 CGAGTGGTGGCAAGATCGCATTTCCAGGTGGCTTGGCCACCGTCAACCAATGCT 1554  
 QY 452 CGGGGTGCGCTACGGCTCCTCTACGAGGATGGTGACAGTGAAGCAGACTTCTGTGCA 511  
 DB 1555 GGGAAAGCGCTTCTCTCGCTACTTCCAT-----TCCGACGACGGCTTTTGGTGTCA 1608  
 QY 512 CATGTGGAGCCCCCTGATCCACCGAGTGGTGGTACGAGCTGGGGCCACGGCAT--- 568  
 DB 1609 CGAGGACTGCGCCCATCTCCATCCAGTGGCTGGGCAACCAAGTAGGCCATAATCTGGC 1668  
 QY 569 -----CAAGATGTACAGAGGCGGAAGTACATGGCCCCATCACCCACCTTCGGAAGA 621  
 DB 1669 TGCACCGCAGGACTATCTGGAGCGCATCTAGCTGGTGGCAAGCCATGATGAGGTTCA 1728  
 QY 622 TCTACTGTGATGCCGTTCCTTACCTCTTCAAGAGGTACGAGCAGTCTACACA----- 674  
 DB 1729 T-GAGGACGATGCCCAATCGAGTGTGTTTAAGATGAATTCACCTTCAGCAATACTACA 1787  
 QY 675 -----GAAGCGGTGGTTTGAAGAGGATGAAGCTGGAGGCCATTTGACCCCTTGA 726  
 DB 1788 GTGACGGCAAAACCAATAGCTTTTGGAGGCGATGAAGCTGGAAGCGGTGGATCCACTCA 1847  
 QY 727 ATCTGGCAACATCTGGTGGCAATGCTGTGAAGTCTCTCTGTGATGATACCTGATGA 786  
 DB 1848 ACCTTTTCATCATATGCCCGGCTACAGTAATGGCGGTTCTTAAGTTCGATACATGATGA 1907  
 QY 787 TCTGTGTGGA---CGGGGGGCGCTCCACAGATGGCTTGGACTGGTCTGTGACCATGCT 843



1908 TACGATTGATTCCTACCAACCGGATGCTCAGGGTCGGATTGGTTCTGTACCATGAAA 1967  
 844 CTTCCAGCCCATCTTCCCGGCCACCTCTCTGTCAGAAAGATGACATGACCTCACACCGC 903  
 1968 AGATCGGTGATCTTCCCGGCTGGATCTCTGTCGTCACCAACATTTCCGTTACCCAC 2027  
 904 CAAAAGGTTATGAGGCACAGACTTCAACTGGGAGAACTACTTGGAGAAAGACCAAGTCGA 963  
 2028 CGAACGGCTACGACTCTGATACATTCACCTGGGAGGGTTACCTCCGCGACACGGAGCG 2087  
 964 AAGCGCTCCATCGAGACTCTTTACATGATGATTCGCCAACCATTGGCTTCAAGTGGGCA 1023  
 2088 TAGCGCTGGCCAGCATCTATTTCATCGGATTTATCCCGATCATGATTTGAGTGGGTA 2147  
 1024 TGAAGCTGGAGCGCGTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACCGTGAAC 1083  
 2148 TGAGTCTGGAGTGTGCAGATCTCTATGATCCCGACTGTTTGGTGCGCCACGGTGGCG 2207  
 1084 GAGTGTGCATCGGCTCTCAGCATTCACATTTGACGGCTGGGACAGGAGTACGACAGT 1143  
 2208 GAGTGGTTGGTCGACTACTCAAGGTTCACTTTGACGGATGAGTACGACCAAGT 2267  
 1144 GGTGTGACTGGAGTCCCGACACATCTACCCGCTCGGTGTGTGAGTCAACGGCTACC 1203  
 2268 GGTGGATTGGATCAGCCGATATATATCCAGTGGATGGTGTGTACTGTGTCAACATA 2327  
 1204 AGCTCCAG---CCTCTGTGCGCGAGAACCGGCCACACCGCTGAAGGCCAAGAGGCCA 1260  
 2328 AGCTAGAGGGCCACCGAGATGACATCAGCAGGCGCCGAAACCGGACCCAAAGGCCA 2387  
 1261 CAAGAAGAAAGAAACAG 1280  
 2388 AAATACGGAAGCGAAAG 2407

RESULT 14  
 ABL19885  
 ID ABL19885 standard; DNA; 4379 BP.  
 AC ABL19885;  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster genomic polynucleotide SEQ ID NO 11128.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX (PEKE) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX WPI; 2001-656860/75.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX Claim 1; SEQ ID NO 11128; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent

capable of detecting 1000 or more genes from Drosophila. The invention is  
 useful in developmental biology and in elucidating cell signalling and  
 cell-cell interactions in higher eukaryotes for the development of  
 insecticides, therapeutics and pharmaceutical drugs. The invention  
 discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 sequences (ABL31840-ABL16175) and the encoded proteins  
 (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 4379 BP; 1162 A; 1151 C; 1118 G; 948 T; 0 other;  
 Query Match 10.4%; Score 266.4; DB 23; Length 4379;  
 Best Local Similarity 55.6%; Pred. No. 5.5e-63;  
 Matches 678; Conservative 0; Mismatches 501; Indels 41; Gaps 7;  
 QY 92 GTACTGGATCGCTCTGTCTATCCAGACAGAGGATCGGTGTCTTCGGTATGAGG 151  
 DB 2091 GTTTTGGGTGGCCACCATCTTGGAAATCAAGGCTATTAAGCCCTAATGACTACGAGG 2150  
 QY 152 CTTTGAATAAGACGCCCATGACTTCTGTGTGCACCTGGGAAACAGTGGATGTCACCC 211  
 DB 2151 TTTTCGATACGGACTC---GCACGACTTCTGGGTGAACCTCTGCAATGCCGAGGTGCATTC 2207  
 QY 212 CATTGCTGTGTGCTCATCAACAGCAAGATCCTAGTGCCTCCACGACCATCCATGCCAA 271  
 DB 2208 GGTGGTGTGTGCGCCATCTCGGGCAAGCCATTAATTCGCGCCGACCATCGACACA 2267  
 QY 272 GTTCCCGACTGGAAAGGCTACCTCATGAACGGCTGTGTGGGCTCCAGAGCTTTCCCGT 331  
 DB 2268 GTACAAGGACTGGAAGGACTTCTGTGTGGAGCTTTATCTCGGAGCCCGACCTTCCCTC 2327  
 QY 332 GGAATTCACATCAGATGGTGGAGCATGAAGTACCCCTTTAGGACGGGATCGCGCT 391  
 DB 2328 CAACTTTTACAACAAAATCAACGACAGCCCTCAGTCGCGCTTCGCGCTGGAATCT 2387  
 QY 392 GGAAGTGTGGACAAAGTCCAGGTGTCCAGCTCGCATGGCTGTGGTGGACACAGTAAT 451  
 DB 2388 CGAGTGGTGGACAAAGGATCGCATTTTCAGGTGGCTGGCCACCGTCACAAAATCGT 2447  
 QY 452 CGGGGTGCGCTACGGCTCTCTACAGAGTGGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 511  
 DB 2448 GGGAAAGCGCTCTTCTGCGCTACTTTCGAT-----TCCGACGACGGCTTTGGTGCA 2501  
 QY 512 CATGTGGAGCCCTCGATCCACCCAGTGGTGTGTCAGAGTGTGGGCGCCAGGAT--- 568  
 DB 2502 CGAGGACTCGCCCATCATCCATCCAGTGGCTGGGCAACCCAGTGAAGCCATATCTGSC 2561  
 QY 569 -----CAAGATGTCAAGAGGCGAAGTGAATGACATGGCCCATCACCACCTTCCGGAAGA 621  
 DB 2562 TGCACCGCAGGACTATCTGGAGCGCATGTAGCTGTGCGGAAAGCCATGATGAGGTTCA 2621  
 QY 622 TCTACTGTGATCGCGTTCCTTACCTCTTCAAGAAAGTACGAGAGTCTACACA----- 674  
 DB 2622 T-GAGAGCATGCCCAATCGAGTTGTTAAGATGAATCTCACCTTCGACGATACTACA 2680  
 QY 675 -----GAAGCGCGTGTGTTTGAAGAGGATGAAGTGAAGTGAAGTGAAGTGAAGTGA 726  
 DB 2681 GTGACGGCAAAACCAATAGCTTTGTGGAGGCGCATGAAGCTGGAAGCGGTGATCCACTCA 2740  
 QY 727 ATCTGGGCAACATCTCGGTGGCACTGTCTGTAAGGTTCTCTCTGATGATGATACCTGATGA 786  
 DB 2741 ACCTTTTCATCATATGCGCGGCTACAGTAATGGCGGTTCTTAAAGTTCGATACATGATGA 2800  
 QY 787 TCTGTGTGA-----CGGGGGGCGCTCCACAGATGGCTTGGACTGTGTGTGTGTGTGTGTGT 843  
 DB 2801 TAGGCATTGATTCCTACCAACCGGATGCTCAGGCTCGGATTTGTTCTGTGTGTGTGTGTGT 2860  
 QY 844 CTTCCACGCCATCTTCCCGGCCACCTTCTGTGCAAGATGACATGAGTGAAGTGAAGTGAAGTGA 903  
 DB 2861 AGAGTCCGTGTATCTTTCGCGGTGATTTCTGTCGTCACCAACCAATTTTCGGTTATCCCGAC 2920



QY 904 CAAAGGTTATGAGCAGACACTTTCACTGGGAGAACTACTTTGGAGAAACCAACTCGA 963  
 Db 2921 CGAACGCTACGACTCTCGTACATTCACTTGGAGGGTTACCTCCGACACGGAGCGG 2980  
 QY 964 AAGCCGCTCATCGAGACTCTTTAAACATGATGTCCTCAACCAATGCTTCAAGTGGGCA 1023  
 Db 2981 TAGCCGCTGCGCAGCATCTATTTCATCGGATTAATCCCGATCATGATTTGAGGTGGTA 3040  
 QY 1024 TGAAGCTGGAGGCGGTGAGCCTGATGAGCCCGGCTCATCTGTGTGGCCACCGTGAAC 1083  
 Db 3041 TGATCTGGAGTGGCAGATCTCATGATCCCGACTCGTTTGGTGGCCACCGTGGCG 3100  
 QY 1084 GAGTGGTCATCGGCTCTCTAGCATCCACTTTGACGCTGGGACAGGATGACAGT 1143  
 Db 3101 GAGTGGTTGGTTCGACTACTCACTCAAGGTTTCACTTTGACGATGGACGATGATGACGACGT 3160  
 QY 1144 GGGTGGACTGCGACTCCCGACACATCTACCCGCTGGCTGTGTGAGCTCACCGGTACC 1203  
 Db 3161 GGTGGATTGGCAATCAGCCGATATATTCAGTCGGATGGTGTACTGTCAACATA 3220  
 QY 1204 AGTCCAG---CCTCTGTGGCCGACAGACCGGCCACACCGCTGAAGGCCAAAGAGGCCA 1260  
 Db 3221 AGCTAGAGGGCCCAACGAGAGTAGCATCAGCAGGCCCCGAAACCGCACCAAGGCCA 3280  
 QY 1261 CAAAGAGAAAAGAAACAG 1280  
 Db 3281 AATACAGCGAAAGCGAAAG 3300

RESULT 15

AAAT23623  
 ID AAAT23623 standard; cDNA to mRNA; 290 BP.  
 AC  
 XX  
 XX  
 DT 05-SEP-1996 (first entry)  
 XX Human gene signature HUMG05481.  
 DE  
 XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
 KW human; cloning; mapping; non-biased library; diagnosis; detection;  
 KW cell typing; abnormal cell function; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FN W09514772-A1.  
 XX  
 PD 01-JUN-1995.  
 XX  
 PF 11-NOV-1994; 94WO-JP01916.  
 XX  
 PR 12-NOV-1993; 93JP-0355504.  
 XX  
 XX (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.  
 XX  
 PI Matsubara K, Okubo K;  
 XX  
 DR WPI; 1995-206931/27.  
 XX  
 XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 XX  
 PS Claim 1; Page 1414; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AAT19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 XX  
 SQ Sequence 290 BP; 46 A; 90 C; 75 G; 64 T; 15 other;

Query Match 9.9%; Score 252.8; DB 16; Length 290;  
 Best Local Similarity 92.7%; Pred. No. 8.6e-60; Indels 1; Gaps 1;  
 Matches 268; Conservative 0; Mismatches 20;  
 QY 1619 GATCCTCTCTGTGTAATAATCTGCCCGTCTGTGAAGCTGGACGCTGGAGACTGTGTG 1678  
 Db 1 GATCCTCTCTGTGTAATAATCTGCCCGTCTGTGAAGCTGGACGCTGGAGACTGTGTG 60  
 QY 1679 GGGTCTCTCTGGACCCGCGCTGTGTTCTGCTCCCTCCCTGTGGAAGTCTATATGACGG 1738  
 Db 61 GGGTCTCTCTGGNACCCGCGCTGTGTTCTGCTCCCTCCCTGTGGAAGTCTATATNACGG 120  
 QY 1739 GCCGCTCTGAGGCCCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTCCCAAGC 1798  
 Db 121 GCTGCTCTGAGGCCCCAGAACTCGTCTGTAAACCACTTTTCCAGCCAGAGTTCCCAAGC 180  
 QY 1799 TGGAACTGCTAGCTGCTCTCTTCTTAAGATGGGCTCCCTCCCGACCGCCCTC 1858  
 Db 181 TGGAACTGCTAGCTGCTCTCTTCTTAAGTGGCTCCCTCCCGACCGCCCTC 240  
 QY 1859 AGTTGCCAGGGATGGGCCA-CCACTGTCACTGTGGAATACAGACA 1906  
 Db 241 AGTTGCCAGGGCTGGGCCAACCCTGTACACTGNGGNTGCANGANA 289

Search completed: February 4, 2004, 13:46:59  
 Job time : 691 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 4, 2004, 13:26:59 ; Search time 9185 Seconds  
(without alignments)  
11379.869 Million cell updates/sec

Title: US-10-031-915-90  
Perfect score: 2555  
Sequence: 1 agacgcccactctatgac.....ttgcttgagaaaaaaaaa 2555

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.un.\*
- 28: em.vi.\*
- 29: em.vl.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rtd.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2555	100.0	2555	6	AX078286	Sequence
2	2553.4	99.9	3194	9	HA305226	AX078286 Homo sapi
3	2553.4	99.9	3222	9	BC017191	AX078286 Homo sapi
4	2538.8	98.4	3188	6	AX0714233	AX0714233 Sequence
5	2538.8	98.4	3188	9	AX056443	AX056443 Homo sapi
6	2438.4	95.4	3299	9	HA305227	AX056443 Homo sapi
7	2046.8	80.1	5875	9	AX074091	AX074091 Homo sapi
8	1434.6	56.1	3071	9	HS0801538	AX074091 Homo sapi
9	1227.8	48.1	2432	9	AX097052	AX097052 Homo sapi
10	1203.6	47.1	3395	10	BC030864	BC030864 Mus muscu
11	1132.4	44.3	8948	9	HS75623	AL035681 Human DNA
12	1128.2	44.2	3399	10	AY237001	AY237001 Mus muscu
13	867.4	33.9	3725	10	AY237002	AY237002 Mus muscu
14	454.8	17.8	458	6	BD115452	BD115452 EST and e
15	413.8	16.2	3590	6	AX046388	AX046388 Sequence
16	410.6	16.1	2878	10	BC020018	BC020018 Mus muscu
17	405.2	15.9	2564	9	AK000062	AK000062 Homo sapi
18	324.4	12.7	400	11	GI4475	GI4475 human STS S
19	281.4	11.0	418	6	AX336757	AX336757 Sequence
20	270.8	10.6	289	11	G43605	G43605 WIAF-2484-S
21	266.4	10.4	4809	3	BT006011	BT006011 Drosophil
22	236.4	9.3	250	11	GI4931	GI4931 human STS S
23	234.8	9.2	164681	2	AC101996	AC101996 Mus muscu
24	210.6	8.2	193660	2	AC102262	AC102262 Mus muscu
25	204.6	8.0	251144	2	AC109565	AC109565 Rattus no
26	204.6	8.0	258549	2	AC133113	AC133113 Rattus no
27	158.8	6.2	77914	3	AC009200	AC009200 Drosophil
28	158.8	6.2	174036	3	AC010214	AC010214 Drosophil
29	158.8	6.2	261775	3	AE003639	AE003639 Drosophil
30	144.6	5.7	199301	2	EX088589	EX088589 Danio rer
31	135.8	5.3	165	6	AX209757	AX209757 Sequence
32	133.6	5.2	2444	9	BC039820	BC039820 Homo sapi
33	133.6	5.2	3394	9	HS089358	HS089358 Homo sapi
34	133.4	5.2	4323	9	AB014581	AB014581 Homo sapi
35	129.2	5.1	62093	2	AC099875	AC099875 Mus muscu
36	127.8	5.0	215435	10	AL662838	AL662838 Mouse DNA
37	124.2	4.9	3315	9	HS0800860	AL110279 Homo sapi
38	122.4	4.8	62093	2	AC099875	AC099875 Mus muscu
39	121.4	4.8	128328	9	AC005839	AC005839 Homo sapi
40	121.4	4.7	612	9	BC012882	BC012882 Homo sapi
41	121.4	4.7	1570	9	BC048421	BC048421 Homo sapi
42	121.4	4.7	2285	9	BC040188	BC040188 Homo sapi
43	117.8	4.6	2636	6	AX156560	AX156560 Sequence
44	102.6	4.0	2231	10	BC036972	BC036972 Mus muscu
45	102.6	4.0	2781	10	AB032165	AB032165 Mus muscu

ALIGNMENTS

RESULT 1	AX078286	AX078286	Sequence 90 from Patent WO0107471.	2555 bp	DNA	linear	PAT 22-FEB-2001
LOCUS	AX078286	Sequence 90 from Patent WO0107471.					
DEFINITION	AX078286	Sequence 90 from Patent WO0107471.					
ACCESSION	AX078286	Sequence 90 from Patent WO0107471.					
VERSION	AX078286.1	GI:13157977					
KEYWORDS		Homo sapiens (human)					
SOURCE		Homo sapiens					
ORGANISM		Homo sapiens					
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE	1	Hillman,J.L., Lal,P., Tang,Y.T., Yue,H., Au-Young,J., Bandman,O.,					
AUTHORS		Azimzai,Y., Yang,J., Lu,D.A., Baughn,M.R., Patterson,C. and Shah,P.					
TITLE		Cell cycle and proliferation proteins					

JOURNAL Patent: WO 0107471-A 90 01-FEB-2001;  
Incyte Genomics, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..2555  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/notes="Incyte ID No: 1868749CB1"  
BASE COUNT 577 a 709 c 722 g 547 t  
ORIGIN  
Query Match 100.0%; Score 2555; DB 6; Length 2555;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2555; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGCAGCTCCACACTCTATGACAGTGGAGAGTGTGATGAAGGGATGAAGTGGAGGTGC 60  
DB 1 AGCAGCTCCACACTCTATGACAGTGGAGAGTGTGATGAAGGGATGAAGTGGAGGTGC 60  
QY 61 TCAACAGTATGTGTGCTCCCGAGCGGGGTGACTGGATCGCCTCTGTTCATCCAGACAG 120  
DB 61 TCAACAGTATGTGTGCTCCCGAGCGGGGTGACTGGATCGCCTCTGTTCATCCAGACAG 120  
QY 121 CAGGGTATCGGGTGTCTCGGTATGAAGGCTTTCAAAATGACGCCAGCCATGACTTCT 180  
DB 121 CAGGGTATCGGGTGTCTCGGTATGAAGGCTTTCAAAATGACGCCAGCCATGACTTCT 180  
QY 181 GGTGCAACCTGGGAACAGTGGATGTCCACCCCATTCGGCTGGTGTGCCATCAACAGCAAGA 240  
DB 181 GGTGCAACCTGGGAACAGTGGATGTCCACCCCATTCGGCTGGTGTGCCATCAACAGCAAGA 240  
QY 241 TCCATGTGCCCCACGGACCATCCATGCCAGTTCACCGACTGGAGGCTACCTCTGTA 300  
DB 241 TCCATGTGCCCCACGGACCATCCATGCCAGTTCACCGACTGGAGGCTACCTCTGTA 300  
QY 301 AACGGCTGTGGGCTCCAGAGCCCTTCCCGTGAATTTCCACATCAAGATGGTGGAGAGCA 360  
DB 301 AACGGCTGTGGGCTCCAGAGCCCTTCCCGTGAATTTCCACATCAAGATGGTGGAGAGCA 360  
QY 361 TGAAGTACCCCTTTAGCAGGGGATCGGGTGGAAAGTGGTGGACAAAGTCCCAAGGTGTAC 420  
DB 361 TGAAGTACCCCTTTAGCAGGGGATCGGGTGGAAAGTGGTGGACAAAGTCCCAAGGTGTAC 420  
QY 421 GCATGCTATGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTACCGCTCTCTACGAGG 480  
DB 421 GCATGCTATGGCTGTGGTGGACACAGTAAATCGGGGTGCGCTACCGCTCTCTACGAGG 480  
QY 481 ATGGTACAGTACGACGACTTCTGGTGCCACATGTGGAGCCCTTGATCCACCCAGTGG 540  
DB 481 ATGGTACAGTACGACGACTTCTGGTGCCACATGTGGAGCCCTTGATCCACCCAGTGG 540  
QY 541 GTTGGTCAACAGCTGTGGGCCACCGCATCAAGATGTCAGAGGCGCAAGTGCATGGCCC 600  
DB 541 GTTGGTCAACAGCTGTGGGCCACCGCATCAAGATGTCAGAGGCGCAAGTGCATGGCCC 600  
QY 601 ATCAACCCCACTTCGGAAGATCTACTGTGATGCCGTTCCCTTACCTCTTCAAGAAGGTAC 660  
DB 601 ATCAACCCCACTTCGGAAGATCTACTGTGATGCCGTTCCCTTACCTCTTCAAGAAGGTAC 660  
QY 661 GAGCAGTCTACAGAAAGGCGGTGGTTGAGAAAGGATGAAGCTGGAGGCCATTGACC 720  
DB 661 GAGCAGTCTACAGAAAGGCGGTGGTTGAGAAAGGATGAAGCTGGAGGCCATTGACC 720  
QY 721 CCCTGAATCTGGGCAACATCTGGTGGCAACTGTCTGTAAAGTTCCTCTGGATGGATACC 780  
DB 721 CCCTGAATCTGGGCAACATCTGGTGGCAACTGTCTGTAAAGTTCCTCTGGATGGATACC 780  
QY 781 TGATGATCTGTGGAGCGGGGCGCTTCCACAGATGCTTGGACTGGTCTGTGCTACCATG 840  
DB 781 TGATGATCTGTGGAGCGGGGCGCTTCCACAGATGCTTGGACTGGTCTGTGCTACCATG 840  
QY 841 CCTCTCCACGCCATCTTCCCGGCCACCTTCTGTGAGAAAGATGATGAGCTCACAC 900  
DB 841 CCTCTCCACGCCATCTTCCCGGCCACCTTCTGTGAGAAAGATGATGAGCTCACAC 900

DB 841 CCTCTCCACGCCATCTTCCCGGCCACCTTCTGTGAGAAAGATGATGAGCTCACAC 900  
QY 901 CGCCAAAGGTTATGAGGCACAGACTTTCAATGGGAGAACTACTTTGGAGAAACCAAGT 960  
DB 901 CGCCAAAGGTTATGAGGCACAGACTTTCAATGGGAGAACTACTTTGGAGAAACCAAGT 960  
QY 961 CGAAAGCCGCTCANTCGAGACTCTTTAAATGATTTGCCAAACCAATGGCTTCAAGGTGG 1020  
DB 961 CGAAAGCCGCTCANTCGAGACTCTTTAAATGATTTGCCAAACCAATGGCTTCAAGGTGG 1020  
QY 1021 GCATGAAGCTGGAGCGCTGAGCTGATGAGAGCCCGGCTCATCTGTGTGGCCACCGTGA 1080  
DB 1021 GCATGAAGCTGGAGCGCTGAGCTGATGAGAGCCCGGCTCATCTGTGTGGCCACCGTGA 1080  
QY 1081 AACAGTGGTGCATCGGCTCTCAGCATCCATTTGAOGGCTGGGACAGGATGACACC 1140  
DB 1081 AACAGTGGTGCATCGGCTCTCAGCATCCATTTGAOGGCTGGGACAGGATGACACC 1140  
QY 1141 AGTGGTGGATCGGAGTCCCGAGACATCTACCCGCTCGGCTGTGTGAGCTCACCGGCT 1200  
DB 1141 AGTGGTGGATCGGAGTCCCGAGACATCTACCCGCTCGGCTGTGTGAGCTCACCGGCT 1200  
QY 1201 ACCAGCTCCAGCCTCTCTGTGGCGGCGAGAACCGGCCACACCGCTGAAAGGCCAAAGAGGCCA 1260  
DB 1201 ACCAGCTCCAGCCTCTCTGTGGCGGCGAGAACCGGCCACACCGCTGAAAGGCCAAAGAGGCCA 1260  
QY 1261 CAAAGAAAGAAAGAAACAGTTTGGGAGAAAGAAAGAAAGTCCCGCCACCTAAGACGC 1320  
DB 1261 CAAAGAAAGAAAGAAACAGTTTGGGAGAAAGAAAGAAAGTCCCGCCACCTAAGACGC 1320  
QY 1321 GACCCCTCAGACAGGGGTCCAAGAGCCCTGTGGAGAGCAGCCCTCAGGGGTGCCAGGA 1380  
DB 1321 GACCCCTCAGACAGGGGTCCAAGAGCCCTGTGGAGAGCAGCCCTCAGGGGTGCCAGGA 1380  
QY 1381 AGATCTCGTGGAGCCTTCTCTGGCGAGATCATTTGCTGTGCGTGTGAAGAGAGAGATC 1440  
DB 1381 AGATCTCGTGGAGCCTTCTCTGGCGAGATCATTTGCTGTGCGTGTGAAGAGAGAGATC 1440  
QY 1441 TAGACGTGGCTCGCCCGACAGAGCTTCAAGTCCAGAGTCCCTCTCGTCCGAGAGAA 1500  
DB 1441 TAGACGTGGCTCGCCCGACAGAGCTTCAAGTCCAGAGTCCCTCTCGTCCGAGAGAA 1500  
QY 1501 TCAAGCAGGAAACAGACGACTGAGCCTTCTGCTGCCAGCTGCTTCTAGCTGGAAGCC 1560  
DB 1501 TCAAGCAGGAAACAGACGACTGAGCCTTCTGCTGCCAGCTGCTTCTAGCTGGAAGCC 1560  
QY 1561 AGCCAGCGGTTCTCTACACCAACCATGCTCCACCTGCTTCCAGCTTGGCTTGGAGACTGA 1620  
DB 1561 AGCCAGCGGTTCTCTACACCAACCATGCTCCACCTGCTTCCAGCTTGGCTTGGAGACTGA 1620  
QY 1621 TCCTCTCTGTGTAATTTCTGCCCGGTCTGTGAAGGCTGGACCGTGGAGACCTGTGGG 1680  
DB 1621 TCCTCTCTGTGTAATTTCTGCCCGGTCTGTGAAGGCTGGACCGTGGAGACCTGTGGG 1680  
QY 1681 GTCTCTGGGAGCCCGCTGTGTCTTCTGCCCTCCCTGTGGAAAGTCTATATGACGGGC 1740  
DB 1681 GTCTCTGGGAGCCCGCTGTGTCTTCTGCCCTCCCTGTGGAAAGTCTATATGACGGGC 1740  
QY 1741 CGCCTGAGGCCCCAGAACTGCTGTGAAACCACTTTTCCAGCCAGAGTTCCCAAGCTG 1800  
DB 1741 CGCCTGAGGCCCCAGAACTGCTGTGAAACCACTTTTCCAGCCAGAGTTCCCAAGCTG 1800  
QY 1801 GAAAGCTAGCTGCTGTCTTCTTAAAGTGGCCTCCCGCCGACCGCCACCGGCCCTCAG 1860  
DB 1801 GAAAGCTAGCTGCTGTCTTCTTAAAGTGGCCTCCCGCCGACCGCCACCGGCCCTCAG 1860  
QY 1861 TTGCCAGGATGGGCGCCACCACTGTCACTGTGGAATACAGACAGTGAACCTGTCTG 1920  
DB 1861 TTGCCAGGATGGGCGCCACCACTGTCACTGTGGAATACAGACAGTGAACCTGTCTG 1920  
QY 1921 CCTGAAACGAGTCAATGTAATTAAGTTCTAGAGCAGCTCTCTGAGCAGGATAAGGTCCCT 1980  
DB 1921 CCTGAAACGAGTCAATGTAATTAAGTTCTAGAGCAGCTCTCTGAGCAGGATAAGGTCCCT 1980

QY	1981	GACAGTCAGTTGTGTGTGTGGGGGAGAGCCTCTGCCTCAAAAATTCAACAAGCAGAAATGCCT	2040
Db	1981	GACAGTCAGTTGTGTGTGTGGGGGAGAGCCTCTGCCTCAAAAATTCAACAAGCAGAAATGCCT	2040
QY	2041	CTCAGCCTCATGTGTTGGTCCTCTGCTCTCTCCTAGCTCCCAAGGATGTTGGGAGCCAG	2100
Db	2041	CTCAGCCTCATGTGTTGGTCCTCTGCTCTCTCCTAGCTCCCAAGGATGTTGGGAGCCAG	2100
QY	2101	CTTGTCTCGCAGCTAAGAAGCAGTGCACAGGATGTGGATTTTGGGCGACTGTGTGGTGG	2160
Db	2101	CTTGTCTCGCAGCTAAGAAGCAGTGCACAGGATGTGGATTTTGGGCGACTGTGTGGTGG	2160
QY	2161	CCTTGAGCTGCTTTCTGTGTTGTGAGGACTGACTCCCATTTCTAAAGGAAATGCCTCC	2220
Db	2161	CCTTGAGCTGCTTTCTGTGTTGTGAGGACTGACTCCCATTTCTAAAGGAAATGCCTCC	2220
QY	2221	GGGAGGACATTTGGGAGGAAGATGGCGTGAAGTGACATTTGGCTCTGTCTACCTGCTCCT	2280
Db	2221	GGGAGGACATTTGGGAGGAAGATGGCGTGAAGTGACATTTGGCTCTGTCTACCTGCTCCT	2280
QY	2281	GAAGCCCCGTAAAAATAATTATCCCAAGATTCTTGTAGTTAAAGGTCAGTTCCTGA	2340
Db	2281	GAAGCCCCGTAAAAATAATTATCCCAAGATTCTTGTAGTTAAAGGTCAGTTCCTGA	2340
QY	2341	CTGGAGCCTCTAGAGAGCTGGGCTGTGATGTTCTTTTGGCTTTTGTCTACCTAAATG	2400
Db	2341	CTGGAGCCTCTAGAGAGCTGGGCTGTGATGTTCTTTTGGCTTTTGTCTACCTAAATG	2400
QY	2401	AAGAAACCATGCTTGGAGGGGCGGTGACACAGNACCCCTCAAGCAAGGATGACAGACT	2460
Db	2401	AAGAAACCATGCTTGGAGGGGCGGTGACACAGNACCCCTCAAGCAAGGATGACAGACT	2460
QY	2461	GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCCAGACTCTGTGTGTGAGA	2520
Db	2461	GGAGGACACATCTAGCTGCCATTGCAACCTCACTGGGCTCCCCAGACTCTGTGTGTGAGA	2520
QY	2521	AATTAAACCCCTGCTTGTGTGAGAAAAA	2555
Db	2521	AATTAAACCCCTGCTTGTGTGAGAAAAA	2555

```

RESULT 2
HSA305226
LOCUS
DEFINITION
HSA305226      3194 bp      mRNA      linear      PRI 18-JAN-2002
Homo sapiens mRNA for H-1(3)mbt-like protein, alternative variant
a.
ACCESSION
AJ305226
VERSION
AJ305226.1      GI:113940238
KEYWORDS
alternative splicing; H-1(3)mbt-like gene; H-1(3)mbt-like protein.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS
Wismar, J.
TITLE
Molecular characterization of h-1(3)mbt-like: a new member of the
human mbt family
JOURNAL
FEBS Lett. 507 (1), 119-121 (2001)
MEDLINE
21538645
PubMed
11692070
REFERENCE
2 (bases 1 to 3194)
AUTHORS
Wismar, J.
TITLE
Direct Submission
JOURNAL
Submitted (25-JAN-2001) Wismar J., Johannes Gutenberg Universitaet,
Institut fuer Genetik, Becherweg 32, 55099 Mainz, GERMANY
FEATURES
Location/Qualifiers
1..3194
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q13.31-33"
source

```

QY 661 GAGCAGTCTACACAGAGGCGGTTGGTTTGGAGAGGAGTGAAGCTGGAGGCCATTGACC 720  
Db 1295 GAGCAGTCTACACAGAGGCGGTTGGTTTGGAGAGGAGTGAAGCTGGAGGCCATTGACC 1354  
QY 721 CCCTGAATCTGGGCAACATCTGGTGGCACTCTCTGTAAGGTTTCCCTGGATGATAC 780  
Db 1355 CCCTGAATCTGGGCAACATCTGGTGGCACTCTCTGTAAGGTTTCCCTGGATGATAC 1414  
QY 781 TGATGATCTGTGGAGCGGGGGCCCTCCACAGATGGCTTGGACTGGTTCTGCTACCATG 840  
Db 1415 TGATGATCTGTGGAGCGGGGGCCCTCCACAGATGGCTTGGACTGGTTCTGCTACCATG 1474  
QY 841 CTTCTTCCACGCGCATCTTCCCGGCGACCTTCTGTGAGAGATGATGAGCTCACAC 900  
Db 1475 CTTCTTCCACGCGCATCTTCCCGGCGACCTTCTGTGAGAGATGATGAGCTCACAC 1534  
QY 901 CGCCAAAAGGTTATGAGGCAAGACTTTCAACTGGGAGAACTACTTGGAGAGAACCAAGT 960  
Db 1535 CGCCAAAAGGTTATGAGGCAAGACTTTCAACTGGGAGAACTACTTGGAGAGAACCAAGT 1594  
QY 961 CGAAAGCCGCTCCATCGAGACTCTTTAAACATGATGATGCGCCAAACCATGCTCAAGTGG 1020  
Db 1595 CGAAAGCCGCTCCATCGAGACTCTTTAAACATGATGATGCGCCAAACCATGCTCAAGTGG 1654  
QY 1021 GCATGAGCTGGAAGCGGCTGAGCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080  
Db 1655 GCATGAGCTGGAAGCGGCTGAGCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1714  
QY 1081 AACGAGTGTGCAATCGGCTCTCTCAGCATCCACTTTGACGCGTGGGACAGCAGTACGACC 1140  
Db 1715 AACGAGTGTGCAATCGGCTCTCTCAGCATCCACTTTGACGCGTGGGACAGCAGTACGACC 1774  
QY 1141 AGTGGTGGAGTGGAGTCCCGAGACATCTACCCGCTCGGCTGGTGGTGGAGCTCACCGCT 1200  
Db 1775 AGTGGTGGAGTGGAGTCCCGAGACATCTACCCGCTCGGCTGGTGGTGGAGCTCACCGCT 1834  
QY 1201 ACCAGCTCCAGCTCTCTGTGGCGCAGAACCGGCGCACACCGCTGGAAGGCGCAAGAGGCCA 1260  
Db 1835 ACCAGCTCCAGCTCTCTGTGGCGCAGAACCGGCGCACACCGCTGGAAGGCGCAAGAGGCCA 1894  
QY 1261 CAAGAAGAAAGAAACAGTTTGGGAGAAAGAAAGAAAGATCCCGGCCCATTAAGACGC 1320  
Db 1895 CAAGAAGAAAGAAACAGTTTGGGAGAAAGAAAGAAAGATCCCGGCCCATTAAGACGC 1954  
QY 1321 GACCCCTCAGACAGGGGTCAGAGAGCCCTGTGTGAGGACGACCTCAGGGGTGCCAGGA 1380  
Db 1955 GACCCCTCAGACAGGGGTCAGAGAGCCCTGTGTGAGGACGACCTCAGGGGTGCCAGGA 2014  
QY 1381 AGATCTCTCGGAGCCTGTTCTGTGGCGAGATCATTTGCTGTGCTGTGAAGGAAGAGCATC 1440  
Db 2015 AGATCTCTCGGAGCCTGTTCTGTGGCGAGATCATTTGCTGTGCTGTGAAGGAAGAGCATC 2074  
QY 1441 TAGAGCTGGCTCGCCGCAAGAGCTTCAAGTCCAGAGCTGCTGTCTCGCTCGAGAACA 1500  
Db 2075 TAGAGCTGGCTCGCCGCAAGAGCTTCAAGTCCAGAGCTGCTGTCTCGCTCGAGAACA 2134  
QY 1501 TCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAGCTCCAGCTTCTAGCTGGAAGCC 1560  
Db 2135 TCAAGCAGGAAACAGACGACTGAGCCTTCTGCTCCAGCTTCTAGCTGGAAGCC 2194  
QY 1561 AGCCAGGTTTCTACACACACACATGCTCCACCTGACTTTGGCTTTGAGACTGA 1620  
Db 2195 AGCCAGGTTTCTACACACACACATGCTCCACCTGACTTTGGCTTTGAGACTGA 2254  
QY 1621 TCCTCTCTGTATAATCTGCGCGGTGTGTGAAGGCTGGACGCTGGAGGACCTGTGGG 1680  
Db 2255 TCCTCTCTGTATAATCTGCGCGGTGTGTGAAGGCTGGACGCTGGAGGACCTGTGGG 2314  
QY 1681 GTCTCTGGGACCGGCTGTGCTTCTGCGCTCCCTGTGGAAAGGTTATATGACGGCC 1740  
Db 2315 GTCTCTGGGACCGGCTGTGCTTCTGCGCTCCCTGTGGAAAGGTTATATGACGGCC 2374  
QY 1741 GCGCTGAGGCGCCAGAACTCGTCTGTGAACCACTTTTCCAGCCAGAGTTTCCCAAGCTG 1800

Db 2375 GCGCTGAGGCGCCAGAACTCGTGTGGAACCACTTTTCCAGCAGAGTTTCCCAAGCTG 2434  
QY 1801 GAAAGCTAGCTGCTGCTCTTCCCTTAAGATGGCTTCCCGCCGAGCCCGCCACGCGCTCAG 1860  
Db 2435 GAAAGCTAGCTGCTGCTCTTCCCTTAAGATGGCTTCCCGCCGAGCCCGCCACGCGCTCAG 2494  
QY 1861 TTGCCAGGATGGGGCCACACATCTGTCACTGTGGAATACAAGACAGTGAATCTGTCTG 1920  
Db 2495 TTGCCAGGATGGGGCCACACATCTGTCACTGTGGAATACAAGACAGTGAATCTGTCTG 2554  
QY 1921 CCTGAAAGAGTCATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATGAAGTCCCT 1980  
Db 2555 CCTGAAAGAGTCATGTAATTAAGTTCTAGACAGCTCTCTGAGCAGGATGAAGTCCCT 2614  
QY 1981 GACAGTGAATGTGTGGGGGAGCCTCTGCTGCTCAAAATTCACCAAGCAGAATGCT 2040  
Db 2615 GACAGTGAATGTGTGGGGGAGCCTCTGCTGCTCAAAATTCACCAAGCAGAATGCT 2674  
QY 2041 CTCAGCCTCATGCTGCTGCTCTCTGCTCTCTAGCTCCCGAGGATGTTGGGAGCCAG 2100  
Db 2675 CTCAGCCTCATGCTGCTGCTCTCTCTCTCTCTAGCTCCCGAGGATGTTGGGAGCCAG 2734  
QY 2101 CTTCTCTCGGAGCTAAAGAGCAGTGAACAGGATGTGGATTTTGGCGACCTGTGTGTGG 2160  
Db 2735 CTTCTCTCGGAGCTAAAGAGCAGTGAACAGGATGTGGATTTTGGCGACCTGTGTGTGG 2794  
QY 2161 CTTGAGCTGCTTCTGTGTTGTGAGGACTGCTCCATTTCTTAAAGGAAATGCCCC 2220  
Db 2795 CTTGAGCTGCTTCTGTGTTGTGAGGACTGCTCCATTTCTTAAAGGAAATGCCCC 2884  
QY 2221 GCGGAGGACATTTGGGAGGAGATGCGCTGAGTGTGCACTTTTGGCTCTGTACCTGCTCT 2280  
Db 2855 GCGGAGGACATTTGGGAGGAGATGCGCTGAGTGTGCACTTTTGGCTCTGTACCTGCTCT 2914  
QY 2281 GAAGCCCGCTTAAATAATTCATCCAAAGTTCCTTTGTAGTTAAAGGTTCCAGTTCTGA 2340  
Db 2915 GAAGCCCGCTTAAATAATTCATCCAAAGTTCCTTTGTAGTTAAAGGTTCCAGTTCTGA 2974  
QY 2341 CTGAGGCTCTAGAGAGTGGGCTGTGTATGTTCTTTTGGCTTTTGTCTTACCTAAATG 2400  
Db 2975 CTGAGGCTCTAGAGAGTGGGCTGTGTATGTTCTTTTGGCTTTTGTCTTACCTAAATG 3034  
QY 2401 AAGAAACCATCCCTGGAGGGCCCTGAACACAGAACCTTCAAGACAGGATGACAGACT 2460  
Db 3035 AAGAAACCATCCCTGGAGGGCCCTGAACACAGAACCTTCAAGACAGGATGACAGACT 3094  
QY 2461 GGAGGACATCTAGCTGCCATTCGAACCTCAGCTGGCTCCCGAGCTCTGTGTGTGAGA 2520  
Db 3095 GGAGGACATCTAGCTGCCATTCGAACCTCAGCTGGCTCCCGAGCTCTGTGTGTGAGA 3154  
QY 2521 AATTAAACCCCTGCTGCTTTGAGAAAAA 2555  
Db 3155 AATTAAACCCCTGCTGCTTTGAGAAAAA 3189

## RESULT 3

BC017191

LOCUS

DEFINITION

Homo sapiens, hypothetical protein DKFp761I141, clone MGC:2476

IMAGE:3138444, mRNA, complete cds.

ACCESSION

BC017191

VERSION

BC017191.1

KEYWORDS

MGC.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 3222)

Straussberg, R.

Direct Submission

TITLE

Submitted (05-NOV-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 Contact: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland;  
 Web site: <http://www.nisac.nih.gov/>  
 Contact: [nisc\\_mgc@hgrl.nih.gov](mailto:nisc_mgc@hgrl.nih.gov)  
 Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q.L., Maslier, C., Maslier, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J., Tongson, E.E., Touchman, J.W., Tsourgeon, C., Vogt, J.L., Walker, M.A., Zhang, J.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 4 Row: b Column: 5  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13899349.

FEATURES

Location/Qualifiers  
 1..3222  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="LocusID:83746"  
 /db\_xref="taxon:9606"  
 /clone="MGC:2476 IMAGE:313844"  
 /tissue\_type="Placenta, choriocarcinoma"  
 /clone\_lib="NIH-MGC\_21"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOTB7"  
 53..2170  
 /codon\_start=1  
 /product="hypothetical protein DKFZp7611141"  
 /protein\_id="AAH17191.1"  
 /db\_xref="GI:16877935"  
 /translation="MEXPRSITEPPSEPMEEEDDLLEFGYDVSFRSYNSVSGES  
 SYLLESEANEENREAGELPTSLHLLSPGTPRDLGSGSEPAVCEMCAIGVTRAF  
 FKTRFCVSCVSYSNKKASILARLOGKPTPKAKVHLKAAASAKGIFLHSG  
 TQADGPTQDQALVLPDQWKEFLKDSYKAPVSCFKHVPYLDQWEDVWKMKEV  
 LNSDAVPSRYWYASVITQAGYRLVLYEGFENDASHDFWNLGTVDVHPICWALN  
 SKILVPRTHIAKFTDMKGLVLMKELVGRSLPVDVDFHIMVSEMKYIPFRGMRLVVDK  
 SQVSRMAVVDVTVIGRLRLYEDGSDDDDFWCHMSPLIHPVGRSRRVHGKIKSE  
 RRSMAHPTPRKTYDAVPLFKKRAVYTEGWFEEGKMLEAIDPLNLGNCVATV  
 CKVLIDGLYIMICVDGPGSTGLDFWFCYHASSHAIFFATPCQKNDIELTPPKYEATF  
 NWENLEKTSKAPSLFNMDCPHGFVKMKLEAVDLMEPRLICVATVRYVHRL  
 SIHEDGSEYDQVDCESDIPYVNGELTGVLOPPVAEPATPLKAEATKKKK  
 QFKRKRIPPTKTRPQKSGSKLEDDPQGAARKISSEVPFGBIIVRVKEHLDVA  
 SPDKASPLPVSVENIKQETDD"  
 BASE COUNT 756 a 870 c 913 g 683 t  
 ORIGIN  
 Query Match 99.9%; Score 2553.4; DB 9; Length 3222;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2554; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 AGCAGTCCCACTTATGACAGTGGGAGGATGTGATGAAAGGATGAAGTGGAGTGC 60  
 DB 648 AGCAGTCCCACTTATGACAGTGGGAGGATGTGATGAAAGGATGAAGTGGAGTGC 707  
 QY 61 TCACAGTGTGCTGTGCTCCCGAGCGGGGTACTGATCGCTCTGTATCCAGACAG 120  
 DB 708 TCACAGTGTGCTGTGCTCCCGAGCGGGGTACTGATCGCTCTGTATCCAGACAG 767  
 QY 121 CAGGATCGGGGTGCTGCTTCGGTATGAAGGCTTTGAAATGACGCCAGCATGACTTCT 180  
 DB 768 CAGGATCGGGGTGCTGCTTCGGTATGAAGGCTTTGAAATGACGCCAGCATGACTTCT 827

QY 181 GGTGCAACCTGGGAACAGTGGATGTCACCCCATTTGGTGGTGTGCCATCAACAGCAAGA 240  
 DB 828 GGTGCAACCTGGGAACAGTGGATGTCACCCCATTTGGTGGTGTGCCATCAACAGCAAGA 887  
 QY 241 TCTAGTGGCCCCCAGGACCATCCATGCCAGTTCACGACGTGGAGGGCTACCTCATGA 300  
 DB 888 TCTAGTGGCCCCCAGGACCATCCATGCCAGTTCACGACGTGGAGGGCTACCTCATGA 947  
 QY 301 AACGGCTGGTGGGCTCCAGAGCGTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 360  
 DB 948 AACGGCTGGTGGGCTCCAGAGCGTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 1007  
 QY 361 TGAAGTACCCCTTTAGGACAGGACATCGGCTGGAAAGTGGTGAAGTCCCAAGTGTCTAC 420  
 DB 1008 TGAAGTACCCCTTTAGGACAGGACATCGGCTGGAAAGTGGTGAAGTCCCAAGTGTCTAC 1067  
 QY 421 GCATCTCGATGGCTGTGGTGGACACAGTAATCGGGGGTCCGCTACGGGCTCTCTACGAGG 480  
 DB 1068 GCATCTCGATGGCTGTGGTGGACACAGTAATCGGGGGTCCGCTACGGGCTCTCTACGAGG 1127  
 QY 481 ATGGTGACAGTGAACGACGACTTCTGTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 540  
 DB 1128 ATGGTGACAGTGAACGACGACTTCTGTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 1187  
 QY 541 GTTGGTCAAGAGTGTGGGCCACGACATCAAGATGTCAGAGAGGCGAAGTGAATGGCC 600  
 DB 1188 GTTGGTCAAGAGTGTGGGCCACGACATCAAGATGTCAGAGAGGCGAAGTGAATGGCC 1247  
 QY 601 ATACCCCCACCTTCCGGGAAGATCTACTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 660  
 DB 1248 ATACCCCCACCTTCCGGGAAGATCTACTGTGATGCGCTTCTTACCTCTTCAAGAGGATAC 1307  
 QY 661 GAGCAGTCTACACAGAGGCGGTTGGTTGAGGAAGGATGAAGTGGAGGCCATTGACC 720  
 DB 1308 GAGCAGTCTACACAGAGGCGGTTGGTTGAGGAAGGATGAAGTGGAGGCCATTGACC 1367  
 QY 721 CCTGAACTCGGGCAACATCTCGGTGGCAACTGTCTGTAAGGTTCTCTGGATGATAC 780  
 DB 1368 CCTGAACTCGGGCAACATCTCGGTGGCAACTGTCTGTAAGGTTCTCTGGATGATAC 1427  
 QY 781 TGATGATCTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGGTTCCTTACCATTG 840  
 DB 1428 TGATGATCTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGGTTCCTTACCATTG 1487  
 QY 841 CCTCTTCCACCGCATCTTCCCGGCCACTTCTGTGAGGAAGTGAATGAGCTCACAC 900  
 DB 1488 CCTCTTCCACCGCATCTTCCCGGCCACTTCTGTGAGGAAGTGAATGAGCTCACAC 1547  
 QY 901 CGCCAAAAGGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTGGAGAGACCAAGT 960  
 DB 1548 CGCCAAAAGGTTATGAGGCACAGACTTTCACTGGGAGAACTACTTGGAGAGACCAAGT 1607  
 QY 961 CGAAAGCGCTCCATCGAGACTCTTTAAACATGATTTGCCCAAAACCATGGCTTCAAGGTGG 1020  
 DB 1608 CGAAAGCGCTCCATCGAGACTCTTTAAACATGATTTGCCCAAAACCATGGCTTCAAGGTGG 1667  
 QY 1021 GCATGAAGCTGGAGCGCTGGACCTGTGAGAGCCCGGCTCATCTGTGTGGCCAGCGTGA 1080  
 DB 1668 GCATGAAGCTGGAGCGCTGGACCTGTGAGAGCCCGGCTCATCTGTGTGGCCAGCGTGA 1727  
 QY 1081 AACGAGTGTGTGATCGGCTCTCAGATTCATTTTGAAGGCTGGGACAGCGAGTAGGACC 1140  
 DB 1728 AACGAGTGTGTGATCGGCTCTCAGATTCATTTTGAAGGCTGGGACAGCGAGTAGGACC 1787  
 QY 1141 AGTGGTGGACTCGGAGTCCCGACAGATCTACCCGCTGGCTGGTGTGAGCTCACCGGCT 1200  
 DB 1788 AGTGGTGGACTCGGAGTCCCGACAGATCTACCCGCTGGCTGGTGTGAGCTCACCGGCT 1847  
 QY 1201 ACCAGTCCAGCTCTCTGTGGCCGAGAACCGGCGCACACCGCTGAAGGCGCAAGAGGCCA 1260  
 DB 1848 ACCAGTCCAGCTCTCTGTGGCCGAGAACCGGCGCACACCGCTGAAGGCGCAAGAGGCCA 1907







1007 TGAAGTACCCCTTTAGGACGGCATCGGCTGGAAGTGGTGGACAACTCCAGGTGTCA 1066  
421 GCACCTCGATGGCTGTGGTGAACACAGTAATCGGGGTGCGCTACGGCTCCTCTACGAGG 480  
1067 GCACCTCGATGGCTGTGGTGAACACAGTAATCGGGGTGCGCTACGGCTCCTCTACGAGG 1126  
481 ATGGTGAAGTGAAGAGTCTTCTGGTGGCCACATGTGGAGCCCGCTGATCCACCCAGTGG 540  
1127 ATGGTGAAGTGAAGAGTCTTCTGGTGGCCACATGTGGAGCCCGCTGATCCACCCAGTGG 1186  
541 GTTGGTCAAGACGTGTGGGCCACCGGCATCAAGATGTCAAGAGCGCAAGTGAATGSCCC 600  
1187 GTTGGTCAAGACGTGTGGGCCACCGGCATCAAGATGTCAAGAGCGCAAGTGAATGSCCC 1246  
601 ATCAACCCACCTTCCGGAAGATCTACTGTGTGATGCGCTTCTTACTCTTCAAGAAGGTAC 660  
1247 ATCAACCCACCTTCCGGAAGATCTACTGTGTGATGCGCTTCTTACTCTTCAAGAAGGTAC 1306  
661 GAGCAGTCTACACGAAGCGGTTGGTTGAGGAAGGATGAAGCTGGAGGCCAATTGACC 720  
1307 GAGCAGTCTACACGAAGCGGTTGGTTGAGGAAGGATGAAGCTGGAGGCCAATTGACC 1366  
721 CCCTGAATCTGGGCAACATCTGCGTGGCACTGTCTGAAGTTCCTCTGATGGATACC 780  
1367 CCCTGAATCTGGGCAACATCTGCGTGGCACTGTCTGAAGTTCCTCTGATGGATACC 1426  
781 TGATGATCTGTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGTACCATG 840  
1427 TGATGATCTGTGTGAGCGGGGGCCCTCCACAGATGGCTTGGACTGTTCTGTACCATG 1486  
841 CCTCTTCCAGCCCATCTTCCGGCCACCTTCTGTCAAGAAATGATGAGTGTACAC 900  
1487 CCTCTTCCAGCCCATCTTCCGGCCACCTTCTGTCAAGAAATGATGAGTGTACAC 1546  
901 CGCCAAAAGGTTATGAGGACAGACTTTCAACTGGGAACTACTTGGAGAAACCAAGT 960  
1547 CGCCAAAAGGTTATGAGGACAGACTTTCAACTGGGAACTACTTGGAGAAACCAAGT 1606  
961 CGAAGCGCTCCATCGAGACTCTTTAAATGATGATTCGCCMAACCATGGCTTCAAGTGG 1020  
1607 CGAAGCGCTCCATCGAGACTCTTTAAATGATGATTCGCCMAACCATGGCTTCAAGTGG 1666  
1021 GCATGAAGCTGGAGCGGTGAGACTGTAGGAGCCCGCTCATCTGTGTGGCCACCGTGA 1080  
1667 GCATGAAGCTGGAGCGGTGAGACTGTAGGAGCCCGCTCATCTGTGTGGCCACCGTGA 1726  
1081 AACGAGTGGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140  
1727 AACGAGTGGTGCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1786  
1141 AGTGGTGAATGCGAGTCCCGAGACATCTACCGGTGCGGTGGTGTGAGCTCACCGCT 1200  
1787 AGTGGTGAATGCGAGTCCCGAGACATCTACCGGTGCGGTGGTGTGAGCTCACCGCT 1846  
1201 ACCAGCTCCAGCTCCTGTGGCCGAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA 1260  
1847 ACCAGCTCCAGCTCCTGTGGCCGAGAACCGGCCACACCGCTGAAGGCCAAAGAGGCCA 1906  
1261 CAAAGAAGAAAGAAACAGTTTGGGAAGAAAGAAAGAAATCCCGCCCACTAAGACGC 1320  
1907 CAAAGAAGAAAGAAACAGTTTGGGAAGAAAGAAAGAAATCCCGCCCACTAAGACGC 1966  
1321 GACCCCTCAGACAGGGGTCCAGAGAGCCCTGTGGAGGACGACCTCAGGGTCCAGGA 1380  
1967 GACCCCTCAGACAGGGGTCCAGAGAGCCCTGTGGAGGACGACCTCAGGGTCCAGGA 2026  
1381 AGATCTGTGGAGCTGTCTTCTGGCGAGATCAATGTCTGTGTGTGAAGGAAGCATC 1440  
2027 AGATCTGTGGAGCTGTCTTCTGGCGAGATCAATGTCTGTGTGTGAAGGAAGCATC 2086  
1441 TAGAGTGGCTCCGCGACAGGCTTCAAGTCCAGAGTGCCTGTCTCGTCCAGACA 1500

2087 TAGAGTGGCTCGCGCCGACAAAGGCTTCAAGTCCAGAGCTCGCTGTCTCCGTGGAACA 2146  
1501 TCAAGCAGGAAACACAGACGACTTCCCTGCTCCAGCTGGCTTCTAGCTGAAGCC 1560  
2147 TCAAGCAGGAAACACAGACGACTTCCCTGCTCCAGCTGGCTTCTAGCTGAAGCC 2206  
1561 AGCCAGGCTTCTCTACACACACCATGCTCCACTGACTTTGGCTTGGAGACTGA 1620  
2207 AGCCAGGCTTCTCTACACACACCATGCTCCACTGACTTTGGCTTGGAGACTGA 2266  
1621 TCCTCTCTGTGTAAATTTCTGCCCGGTGTGTGAAGGCTGGAGCTGGAGGACTCTCTGGG 1680  
2267 TCCTCTCTGTGTAAATTTCTGCCCGGTGTGTGAAGGCTGGAGCTGGAGGACTCTCTGGG 2326  
1681 GTCTCTGGGACCCCGCTTGTCTTCTGCCCTTCCCTTGAAGAAAGTCTATATGACGGG 1740  
2327 GTCTCTGGGACCCCGCTTGTCTTCTGCCCTTCCCTTGAAGAAAGTCTATATGACGGG 2386  
1741 CGCCTGAGGCCCCAGAACTCGTCTGTGAACACCTTTTCCAGCCAGAGTCCCAAGCTG 1800  
2387 CGCCTGAGGCCCCAGAACTCGTCTGTGAACACCTTTTCCAGCCAGAGTCCCAAGCTG 2446  
1801 GAACGCTAGCTGCTGCTCTTCTTAAAGATGGCTTCCCGGACCCCGGACCGGCTCTCAG 1860  
2447 GAACGCTAGCTGCTGCTCTTCTTAAAGATGGCTTCCCGGACCCCGGACCGGCTCTCAG 2506  
1861 TTGCCAGGGATGGGCGACACACTGTGTGAATACAAGACAGTGAATCTGTCTG 1920  
2507 TTGCCAGGGATGGGCGACACACTGTGTGAATACAAGACAGTGAATCTGTCTG 2566  
1921 CTTGAACGAGTCACTGTAAATTAAGTCTAGACAGCTCTCTGAGCAGGATGAAGTCCCC 1980  
2567 CTTGAACGAGTCACTGTAAATTAAGTCTAGACAGCTCTCTGAGCAGGATGAAGTCCCC 2626  
1981 GACAGTGAATGTGTGGTGGGCGAGCTCTGCTTCAAAAATTCACCAAGCAGATGCT 2040  
2627 GACAGTGAATGTGTGGTGGGCGAGCTCTGCTTCAAAAATTCACCAAGCAGATGCT 2686  
2041 CTCAGCTCATGTGTGTGCT 2100  
2687 CTCAGCTCATGTGTGTGCT 2746  
2101 CTTGTCTCGGACAGCTAAGAACGAGTGAACAGGATGTGATTTGGGACCTGTGTGGTGG 2160  
2747 CTTGTCTCGGACAGCTAAGAACGAGTGAACAGGATGTGATTTGGGACCTGTGTGGTGG 2806  
2161 CTTGAGCTGCTTCTGT 2220  
2807 CTTGAGCTGCTTCTGT 2866  
2221 GGGGAGGACATTTGGGAGGAGATGGCTGAGTGTGCACTTTTGGCTCTGTCTCTCTCTCT 2280  
2867 GGGGAGGACATTTGGGAGGAGATGGCTGAGTGTGCACTTTTGGCTCTGTCTCTCTCTCT 2926  
2281 GAAGCCCCCTAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2340  
2927 GAAGCCCCCTAAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2986  
2341 CTGGAGCTCTAGAGCTGGGCTTGTATGTCTTTTGGCTTTTGTCTCTCTCTCTCTCTCT 2400  
2987 TTGGAGCTCTAGAGCTGGGCTTGTATGTCTTTTGGCTTTTGTCTCTCTCTCTCTCTCT 3046  
2401 AAGAAACCATGTCTGGAGGGGCGTGAACACAGAACCCCTCAAGACAGGATGACAGCT 2460  
3047 AAGAAACCATGTCTGGAGGGGCGTGAACACAGAACCCCTCAAGACAGGATGACAGCT 3106  
2461 GAGGACACATCTAGCTGCATGCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2520  
3107 GAGGACACATCTAGCTGCATGCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3166  
2521 AATTAACCCCTGCTTGTCTG 2542  
3167 AATTAACCCCTGCTTGTCTG 3188

```

RESULT 5
AK056443
LOCUS
DEFINITION Homo sapiens cDNA FLJ1881 fis, clone NT2RP7002829, weakly similar
to Scn-related gene containing four mbt domains.
ACCESSION AK056443
VERSION 1
KEYWORDS oligo capping, fis (full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
1 Unpublished
2 (bases 1 to 3188)
REFERENCE Isogai,T., Otsuki,T. and Sugiyama,T.
AUTHORS Direct Submission
TITLE Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
JOURNAL Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(BE-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- and 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES
source
1. 3188
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RP7002829"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2RP7"
/notes="cloning vector: pME18SFL3-mRNA from NT2 neuronal
precursor cells after 5-weeks retinoic acid (RA)
induction."
BASE COUNT 725 a 869 c 909 g 685 t
ORIGIN
Query Match 99.4%; Score 2538.8; DB 9; Length 3188;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2540; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGCAGTCCCACTTATGACCAAGTGGAGGATGTGATGAAGAGGATGAAGTGGAGGTGC 60
DB 647 AGCAGTCCCACTTATGACCAAGTGGAGGATGTGATGAAGAGGATGAAGTGGAGGTGC 706
QY 61 TCACAGTGTGCTGCTCCCAAGCGGGTGTACTCGCTCGCTCTGTCATCCAGACAG 120
DB 707 TCACAGTGTGCTGCTGCTCCCAAGCGGGTGTACTCGCTCTGTCATCCAGACAG 766
QY 121 CAGGATTCGGGTGCTGCTCGGTATGAAGCTTTGAAATGACGCGACCATGCTTCT 180
DB 767 CAGGATTCGGGTGCTGCTCGGTATGAAGCTTTGAAATGACGCGACCATGCTTCT 826
QY 181 GGTGCACTTGGGACAGTGTGATGCCACCCATTTGGCTGGTGGCTCAACAGCAAGA 240
DB 827 GGTGCACTTGGGACAGTGTGATGCCACCCATTTGGCTGGTGGCTCAACAGCAAGA 896

```

```

QY 241 TCCTAGTGCCTCCACGACCATCCATGCCAGTTTCCCGCTGGAAGGGCTACCTCATGA 300
DB 887 TCCTAGTGCCTCCACGACCATCCATGCCAGTTTCCCGCTGGAAGGGCTACCTCATGA 946
QY 301 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 360
DB 947 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGGATTTCCACATCAAGATGGTGGAGACA 1006
QY 361 TGAAGTACCCCTTTAGGACAGGGCATCGGCTGGAAGTGGTGGCAAGTCCCAGAGTGCAC 420
DB 1007 TGAAGTACCCCTTTAGGACAGGGCATCGGCTGGAAGTGGTGGCAAGTCCCAGAGTGCAC 1066
QY 421 GCATCTCGCATGGCTGTGTGGACACAGTAATCGGGGGTGGCTACGGCTCTCTACGAGG 480
DB 1067 GCATCTCGCATGGCTGTGTGGACACAGTAATCGGGGGTGGCTACGGCTCTCTACGAGG 1126
QY 481 ATGGTGACAGTGAAGACGACTTCTGGTGGCCATGTGGAGCCCCCTGATCCACCCAGTGG 540
DB 1127 ATGGTGACAGTGAAGACGACTTCTGGTGGCCATGTGGAGCCCCCTGATCCACCCAGTGG 1186
QY 541 GTTGTGTACAGCGTGTGGGCCACCGCATCAAGATGTGAGAGGCGAAGTGACATGGCCCC 600
DB 1187 GTTGTGTACAGCGTGTGGGCCACCGCATCAAGATGTGAGAGGCGAAGTGACATGGCCCC 1246
QY 601 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCTTACTCTTCAAGAGGTAC 660
DB 1247 ATCAACCCCACTTCCGGAAGATCTACTGTGATGCCGTTCTTACTCTTCAAGAGGTAC 1306
QY 661 GAGCAGTCTACACAGAGAGCGGTTGTTTGAAGAGGATGAAGCTGGAGGCCATTGACC 720
DB 1307 GAGCAGTCTACACAGAGAGCGGTTGTTTGAAGAGGATGAAGCTGGAGGCCATTGACC 1366
QY 721 CCCTGAATCTGGGCAACATCTGCGTGGCAACTGTCTGTAAAGTTCTCTCGATGGATACC 780
DB 1367 CCCTGAATCTGGGCAACATCTGCGTGGCAACTGTCTGTAAAGTTCTCTCGATGGATACC 1426
QY 781 TGATGATCTGTGGAGCGGGGGCCCTCCACAGATGGCTTGGATGGTCTGTGCTACCATG 840
DB 1427 TGATGATCTGTGTGGAGCGGGGGCCCTCCACAGATGGCTTGGATGGTCTGTGCTACCATG 1486
QY 841 CCTCTTCCACGCCATCTTCCCGGCCACCTTCTCTCGAAGATGACATTGAGTCTACAC 900
DB 1487 CCTCTTCCACGCCATCTTCCCGGCCACCTTCTCTCGAAGATGACATTGAGTCTACAC 1546
QY 901 CGCCAAAAGGTTATGAGGCACAGACTTCAACTGGGAGAACTACTTGGAGAGACCAAGT 960
DB 1547 CGCCAAAAGGTTATGAGGCACAGACTTCAACTGGGAGAACTACTTGGAGAGACCAAGT 1606
QY 961 CGAAAGCGCTCCCATCGAGACTCTTTAACTGGATTGCCCAAAACCATGGCTTCAAGTGG 1020
DB 1607 CGAAAGCGCTCCCATCGAGACTCTTTAACTGGATTGCCCAAAACCATGGCTTCAAGTGG 1666
QY 1021 GCATGAAGCTGGAGCGCGTGACCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
DB 1667 GCATGAAGCTGGAGCGCGTGACCTGATGAGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1726
QY 1081 AACGAGTGGTCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1140
DB 1727 AACGAGTGGTCATCGGCTCCTCAGCATCCACTTTGACGGCTGGGACAGCGAGTACGACC 1786
QY 1141 AGTGGGTGGATCGGAGTCCCACAGACATCTACCCCGTGGGTGGTGTGAGTCAACGGCT 1200
DB 1787 AGTGGGTGGATCGGAGTCCCACAGACATCTACCCCGTGGGTGGTGTGAGTCAACGGCT 1846
QY 1201 ACCAGTCCAGCTTCTGTGGCCGAGNACCGGCGACACCGCTGAAGGCCAAGAGGCCA 1260
DB 1847 ACCAGTCCAGCTTCTGTGGCCGAGNACCGGCGACACCGCTGAAGGCCAAGAGGCCA 1906
QY 1261 CAAGAAGAAAAAGAACAGTTTGGGAAGAAAAAGAAAAAGAAATCCCGCCCACTAAGACGC 1320
DB 1907 CAAGAAGAAAAAGAACAGTTTGGGAAGAAAAAGAAAAAGAAATCCCGCCCACTAAGACGC 1966

```

1321	QY	GA	CCCTCTGACAGAGGGGTCAAGAAAGCCCTGCTGGAGAGACGACCCCTCAAGGGTGCACGGA	1380
1967	DB	GA	CCCTCTCAGACAGGGGTCCAAGAAAGCCCTGCTGGAGAGACGACCCCTCAAGGGTGCACGGA	2026
1381	QY	AGATCTGCTCGGAGCCGTGTTCCCTGGCGAGATCATTTGTCGTCGTGTAAGAGAGAGCATC	1440	
2027	DB	AGATCTGCTCGGAGCCGTGTTCCCTGGCGAGATCATTTGTCGTCGTGTAAGAGAGAGCATC	2086	
1441	QY	TAGACGTGGCCCTCGCCCGACAAGAGCTTCAAGTCCAGAGCTGCCTGCTCTCCGTGAGAAACA	1500	
2087	DB	TAGACGTGGCCCTCGCCCGACAAGAGCTTCAAGTCCAGAGCTGCCTGCTCTCCGTGAGAAACA	2146	
1501	QY	TCAGAGCAGGAAAACAGACGACTGAGCCCTTCCTGCTCCAGCCTGCCTTCTAGCTGGAAGCC	1560	
2147	DB	TCAGAGCAGGAAAACAGACGACTGAGCCCTTCCTGCTCCAGCCTGCCTTCTAGCTGGAAGCC	2206	
1561	QY	AGCCACGCGTTTCTCTACCAACCACCAACCATGCTCCACCTGACGCTTTGGGCTTGGAGACTGA	1620	
2207	DB	AGCCACGCGTTTCTCTACCAACCACCAACCATGCTCCACCTGACGCTTTGGGCTTGGAGACTGA	2266	
1621	QY	TCCTCTCTGTGTAAATTCCTGCGCGGTCTGTGAAGCTTGACGGTGGAGGACCTGCTGGG	1680	
2267	DB	TCCTCTCTGTGTAAATTCCTGCGCGGTCTGTGAAGCTTGACGGTGGAGGACCTGCTGGG	2326	
1681	QY	GTCTCTCTGGGACCGCCCTGTTGCTTCTGECCTCCCTCTGTGAAAGGCTATATGACGGGC	1740	
2327	DB	GTCTCTCTGGGACCGCCCTGTTGCTTCTGECCTCCCTCTGTGAAAGGCTATATGACGGGC	2386	
1741	QY	CGCTGAGGCCCCAGAACTCGTCTGTGAAACCACTTTTTCAGCCAGAGTGTCCCAAAGCTG	1800	
2387	DB	CGCTGAGGCCCCAGAACTCGTCTGTGAAACCACTTTTTCAGCCAGAGTGTCCCAAAGCTG	2446	
1801	QY	GAACGCTAGCTGCTGCTCTTCTTAAAGATGGCCTCCCGCCGACCCGCCACGGCCCTCAG	1860	
2447	DB	GAACGCTAGCTGCTGCTCTTCTTAAAGATGGCCTCCCGCCGACCCGCCACGGCCCTCAG	2506	
1861	QY	TTGCCAGGGATGGGCCCAACACTGTCACTGTGGAAATACAAGACAGTGAACCTCTGTCTG	1920	
2507	DB	TTGCCAGGGATGGGCCCAACACTGTCACTGTGGAAATACAAGACAGTGAACCTCTGTCTG	2566	
1921	QY	CCTGAACGAGTCAATGTAAATTAAGTTCTTAGACAGCTCTCTGACAGGATAAGTCCCT	1980	
2567	DB	CCTGAACGAGTCAATGTAAATTAAGTTCTTAGACAGCTCTCTGACAGGATAAGTCCCT	2626	
1981	QY	GACAGTGAGTTGTGTGTGGGGCAGGCTCTGCCCTCAAAATTCACCAAGCAGAAATGCCT	2040	
2627	DB	GACAGTGAGTTGTGTGTGGGGCAGGCTCTGCCCTCAAAATTCACCAAGCAGAAATGCCT	2686	
2041	QY	CTCAGCCCTCATGTGTTGGTCCCTCTGCTCTCTAGCTCCCGCAGGGATTTGGGGACCCAG	2100	
2687	DB	CTCAGCCCTCATGTGTTGGTCCCTCTGCTCTCTAGCTCCCGCAGGGATTTGGGGACCCAG	2746	
2101	QY	CTTGCTCTGGCAGCTAAGAAAGCAGTGAACCAAGATGTGGAATTTTGGCCACCTGTGTGGTG	2160	
2747	DB	CTTGCTCTGGCAGCTAAGAAAGCAGTGAACCAAGATGTGGAATTTTGGCCACCTGTGTGGTG	2806	
2161	QY	CTTGAAGCTGCTTTCTGTGTTTGGAGACTGACTCCCAATTTCTAAAGGAAATGCCCC	2220	
2807	DB	CTTGAAGCTGCTTTCTGTGTTTGGAGACTGACTCCCAATTTCTAAAGGAAATGCCCC	2866	
2221	QY	GGGAGAGACATTTGGAGGAAAGATGGCCTGAGTGTGCACTTTTGGCTCTGCTACCTGCTCCT	2280	
2867	DB	GGGAGAGACATTTGGAGGAAAGATGGCCTGAGTGTGCACTTTTGGCTCTGCTACCTGCTCCT	2926	
2281	QY	GAAGCCCCGCTAAAAATAATTCATCCAAGATTCCTTTGTAGTTAAAGGGTCCAGTCTCTGA	2340	
2927	DB	GAAGCCCCGCTAAAAATAATTCATCCAAGATTCCTTTGTAGTTAAAGGGTCCAGTCTCTGA	2986	
2341	QY	CTGAGGCTCTAGAGAGCTGGGCTGTATGTTCTTTTGGCTTTTGTCTCTACTCAATATG	2400	
2987	DB	TTGAGGCTCTAGAGAGCTGGGCTGTATGTTCTTTTGGCTTTTGTCTCTACTCAATATG	3046	
2401	QY	AAGAAACCATGCTCTGGAGGGGCCGTGAAACACAGAACCCCTCAAGACAAGGATGACAGAGCT	2460	

Db 3047 AGAAGAACCATGCTGAGGGGCGTGAAACAGAACCCCTCAACACAGGATGACAGAGCT 3106  
 QY 2461 GGAGGACACATCTAGCTGCAATTCACCACTCTCTGGGCTCCCCAGAGCTCTGTGTGTGAGA 2520  
 Db 3107 GGAGGACACATCTAGCTGCAATTCACCACTCTCTGGGCTCCCCAGAGCTCTGTGTGTGAGA 3166  
 QY 2521 AATTAACCCCTGCTGCTGCTG 2542  
 Db 3167 AATTAACCCCTGCTGCTGCTG 3188  
 RESULT 6  
 HSA305227  
 LOCUS  
 DEFINITION Homo sapiens mRNA for H-1(3)mbt-like protein, alternative variant b.  
 ACCESSION AJ305227  
 VERSION AJ305227.1 GI:13940240  
 KEYWORDS alternative splicing; H-1(3)mbt-like gene; H-1(3)mbt-like protein.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Wismar, J.  
 TITLE Molecular characterization of h-1(3)mbt-like: a new member of the human mbt family  
 JOURNAL FEBS Lett. 507 (1), 119-121 (2001)  
 MEDLINE 21538645  
 PUBMED 11682070  
 REFERENCE 2 (bases 1 to 3299)  
 AUTHORS Wismar, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-JAN-2001) Wismar J., Johannes Gutenberg Universitaet, Institut fuer Genetik, Becherweg 32, 55099 Mainz, GERMANY  
 FEATURES  
 source  
 1..3299  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /map="22q13.31-33"  
 1..3299  
 /gene="H-1(3)mbt-like"  
 40..1884  
 /gene="H-1(3)mbt-like"  
 /note="alternative variant b"  
 /codon\_start=1  
 /product="H-1(3)mbt-like protein"  
 /protein\_id="CAC37795.1"  
 /db\_xref="GI:13940241"  
 /translation="MEKPSRIETPTSPSEPEEEDDLLEFGGYDPSRYSNVSYSVSSYSSLESAENEDRAGBLPTSLHLLSGTSPRLSDGSGSEPAVCEMCGIVGTREAF FSKTEFCSVCSRSYSSNKASILARLQKGPPTKAKVLHKAWSAKIGAFHLSQQTQLADGTTGQDALVLGPDWGFGLKDHYSKAPVSCFKHVLVDQEDVWGMKVSV LNSDAVPRVYIASVIQTGRVILLRYEGFENDASHDFWNLGTFDVHPHIGNCALIN SKTLVPPRTIAKFTDWKGLYMLRVGSRITLPDFHIKVESMKIPFQGRRLVEVDK SSVSRTRMAVDVTICGRLLRYEDGSDDDDFWCHWSPSLIHPVGRSRRVGHGIMSEI RSDMAHPTFRKIYCDAPVYLFKVRVYTEGWFEEGMKLEAIDPLNGTICVATF KVLLDGLYIM CVDGPGTDLGDFCFYASHASHAIFPATFCQKNDIELTPKGYEAFQF NNEVLEKTSKAPSRLLNMDCPNHFVKYGMKLEAVDLMEPRLICVATVKRVVHRL SIFHDGEYEDWDCESDPIYIPYWCMLTGVLQPPVAAGKLPRSL"  
 BASE COUNT 758 a 909 c 932 g 700 t  
 ORIGIN  
 Query Match 95.4%; Score 2438.4; DB 9; Length 3299;  
 Best Local Similarity 96.0%; Pred. No. 0;  
 Matches 2554; Conservative 0; Mismatches 1; Indels 105; Gaps 1;  
 QY 1 AGCAAGTCCCACTCTATGACCACTGGAGGATGATGATGAACGATGAACGATGCGGTGC 60  
 Db 635 AGCAAGTCCCACTCTATGACCACTGGAGGATGATGATGAACGATGAACGATGCGGTGC 694











```
Db 5796 AGGCACATCTAGTCGCATGACCACTCTACTGGCTCCCGACACTCTGTGTGAGAAA 5855
|||||
Qy 2523 TTAACCCCTGCTGCTTG 2542
|||||
Db 5856 TTAACCCCTGCTGCTTG 5875
|||||

RESULT 8
HSM801538 3071 bp mRNA linear PRI 20-MAR-2002
LOCUS Homo sapiens mRNA; cDNA DKFZp7611141 (from clone DKFZp7611141);
DEFINITION complete cds.
ACCESSION AL136564
VERSION AL136564.1 GI:13276634
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3071)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J. and
Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-2002) MIPS, Am Klopferspitze 18a, D-82152
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. This clone
(DKFZp761141) is available at the RZPD in Berlin. Please contact
the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://mips.gsf.de/proj/cDNA/.
FEATURES
source
1..3071
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="22q13.31-13.33"
/clones="DKFZp7611141"
/tissue_type="amygdala"
/clone_lib="761 (synonym: hamy2). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
1..3071
/gene="DKFZp7611141"
55..1908
/gene="DKFZp7611141"
/translation="similarity to Human 1(3)mbt protein homolog mRNA"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB66499.1"
/db_xref="GI:13276635"
/db_xref="SPTREMBL:O9BO12"
/translation="MEKPSRIEETPSSEPMEEEDDDLELPGYDSFRSYNSVGSSES
SYLEESSEANEDEAGELPTPLHLSPTPSRLDGSSEFAVCEMCGIVGTREAF
FSKTRCSVSRSYSSNKKASILARLOQPTPKAKVLHKAWSAKIGAFILHSQV
TQLADGTPQDQALVGFDMGKFLDHSYKAAFPVSCFKEVFLYDQWEDYMGKMYEV
LNSDAVLPSRYVLAISYQTAGYRVLRYEGFENDASHDFWNCNLGTVDPHPIGWCAIN
SKILVPRPIHAKTDKNGYLMKRLVGRSLPVDPHI.KMVESMKYPPROGMRLVVDK
SOVSRVMAVDTVIGLRLLYEDGSDDDFWCHMSPLIHPVGNRSRVGHGIMKSE
RSDMAHPTFRKYLCAVPLYPKKRVAVTEGWFEGMKLEALDPLNLGNICVAVI
CKVLIDGLYMLICVDDGSPDGLDFWCFHASSHALFPATFCQKNDIEUTPPAGYEAQTF
NWENYLEKTSKAPAFKFNMDPCNHFVKGMKLEAVLMEPLRICVATVTKRVVHRL
SIHFDGMDSEYDQWDCESDPIYPVGMCELTGTQLOPPVAAGVSGSRGPKRL"
3052
polya_site /gene="DKFZp7611141"
665 a 882 c 958 g 665 t
BASE COUNT 56.1%; Score 1434.6; DB 9; Length 3071;
ORIGIN
Query Match
```

```
Best Local Similarity 79.8%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 9; Indels 480; Gaps 1;
Qy 1 AGCAGTCCCACTCTATGACACAGTGGGAGGATGTGTAAGAGGATGAAGTGGAGGTGC 60
Db 650 AGCAGTCCCACTCTATGACACAGTGGGAGGATGTGTAAGAGGATGAAGTGGAGGTGC 709
Qy 61 TCAACAGTGATGCTGTGCTCCCGACGCGGGGTACTGGATCGCCTCTGTCTCATCAGACAG 120
Db 710 TCAACAGTGATGCTGTGCTCCCGACGCGGGGTACTGGATCGCCTCTGTCTCATCAGACAG 769
Qy 121 CAGGGTATCGGGTGTGCTTCGGTATGAAGCTTTGAAATAGCCAGCCATGACTTCT 180
Db 770 CAGGGTATCGGGTGTGCTTCGGTATGAAGCTTTGAAATAGCCAGCCATGACTTCT 829
Qy 181 GGTCAACCTCGGACACAGTGTATGCCACCCATTGGCTGGTGTGCCATCAACAGCAAGA 240
Db 830 GGTCAACCTCGGACACAGTGTATGCCACCCATTGGCTGGTGTGCCATCAACAGCAAGA 889
Qy 241 TCTAGTGCCTCCCGACGACCATCCATGCCAAGTTCACCGACTGGAAGGGTACCTCATGA 300
Db 890 TCTAGTGCCTCCCGACGACCATCCATGCCAAGTTCACCGACTGGAAGGGTACCTCATGA 949
Qy 301 AACGCTGTGGGCTCCAGGACGCTTCCGCTGGATTCCACATCAAGATGCTGGAGACA 360
Db 950 AACGCTGTGGGCTCCAGGACGCTTCCGCTGGATTCCACATCAAGATGCTGGAGACA 1009
Qy 361 TGAAGTACCCCTTTAGGCAGGCGCATGCCGCTGGAAGTGGTGGACAAAGTCCAGAGTGTAC 420
Db 1010 TGAAGTACCCCTTTAGGCAGGCGCATGCCGCTGGAAGTGGTGGACAAAGTCCAGAGTGTAC 1069
Qy 421 GCATCGCATGCTGTGTGACACAGTAATCGGGGTGCGCTACGGTCTCTTACGAGG 480
Db 1070 GCATCGCATGCTGTGTGACACAGTAATCGGGGTGCGCTACGGTCTCTTACGAGG 1129
Qy 481 ATGGTGACAGTGAACGACGACTTCTGCTGCCACATGTGGAGCCCCCTGATCCACCCAGTGG 540
Db 1130 ATGGTGACAGTGAACGACGACTTCTGCTGCCACATGTGGAGCCCCCTGATCCACCCAGTGG 1189
Qy 541 GTTGTGACAGCTGTGGGCCACGCGCATCAAGATGTCTAGAGAGGCGAAGTGAACATGGGCC 600
Db 1190 GTTGTGACAGCTGTGGGCCACGCGCATCAAGATGTCTAGAGAGGCGAAGTGAACATGGGCC 1249
Qy 601 ATCAACCCACCTTCGCGAGAGATCTACTGTGATCGCGTTCCTTACCTCTTCAAGAGGTAC 660
Db 1250 ATCAACCCACCTTCGCGAGAGATCTACTGTGATCGCGTTCCTTACCTCTTCAAGAGGTAC 1309
Qy 661 GAGCAGTCTACACAGAGCGGCTTGGTTTGAAGAGGAGATGAAGTGGAGGCCATTGACC 720
Db 1310 GAGCAGTCTACACAGAGCGGCTTGGTTTGAAGAGGAGATGAAGTGGAGGCCATTGACC 1369
Qy 721 CCTGAAATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGTTCTCTGGATGGATACC 780
Db 1370 CCTGAAATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGTTCTCTGGATGGATACC 1429
Qy 781 TGATGATCTGTGTGACGCGGGGCGCTCCACAGATGGCTTGGACTGGTTCTGCTACCATG 840
Db 1430 TGATGATCTGTGTGACGCGGGGCGCTCCACAGATGGCTTGGACTGGTTCTGCTACCATG 1489
Qy 841 CTTCTTCCACGCCCATCTTCCCGGCACTTGTCTGACAGAAATGACATTTAGCTCACAC 900
Db 1490 CTTCTTCCACGCCCATCTTCCCGGCACTTGTCTGACAGAAATGACATTTAGCTCACAC 1549
Qy 901 CGCCAAAAGTTATGAGGCACAGACTTTTCACTGGGAGAACTACTTGGAGAGAGACCAAGT 960
Db 1550 CGCCAAAAGTTATGAGGCACAGACTTTTCACTGGGAGAACTACTTGGAGAGAGACCAAGT 1609
Qy 961 CGAAAAGCGCTCCATCGAGACTCTTTAAATCATGGATTTGCCAAACCATGGCTTCAAGGTG 1020
Db 1610 CGAAAAGCGCTCCATCGAGACTCTTTAAATCATGGATTTGCCAAACCATGGCTTCAAGGTG 1669
Qy 1021 GATGAGCTGAGGCGCGGTGGACCTGATGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080
```



PTKAKVHLKAAWSAKIGAFHLSQGTQOLADGTPGTQDALVLPDNGKFLKDRHSYKAA  
PVSCFKHVLPLVDQEDVNMKMEVNSDAVLPSRVYVIAASVITAGTVLLRYEGFE  
NDASDFWNLGTVDVHPPIGWCAINSKILVPTPIHAKFTDMKGLMKRLVGRSTLPV  
DFHUKWBSMKVPRQGRWLEVDKQSVSRTRMAVVDVTVIGRLRLLYEDGSDDDDFW  
CHMSPLPHVWMSRVCHGIGKMSERSDMAHPTFKIYCDAPVPLPKVRVAVYEG  
GWFEGMKLEALDPLNLGNIYATVCKVLLDGYLMIYDVGPSFDGLDWFCHASSEA  
IFPATFCOKNDIEPTTPKGYEAFINWENILEKTKSAAPSRLEFNMDCPNHFKVGMK  
LEADLMFPRICVATKRVVHRLLSIHFQDWDSEYDQWVDCSPDIPYVGVCELTGY  
QLQPPVAAVGSRGPKEL"

BASE COUNT 504 a 682 c 696 g 550 t

ORIGIN

Query Match 48.1%; Score 1227.8; DB 9; Length 2432;  
Best Local Similarity 99.4%; Pred. No. 1.1e-303;  
Matches 1232; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AGCAGTCCCACTCATGACCAAGTGGGAGGATGTGATGAAGGAGTGAAGTGGAGGTGC 60  
DB 719 AGCAGTCCCACTCATGACCAAGTGGGAGGATGTGATGAAGGAGTGAAGTGGAGGTGC 778  
QY 61 TCAACAGTGTGCTGTGCTCCCGAGCGGGTGTACTGTGATCGCTCTGTGATCCAGACAG 120  
DB 779 TCAACAGTGTGCTGTGCTCCCGAGCGGGTGTACTGTGATCGCTCTGTGATCCAGACAG 838  
QY 121 CAGGATATCGGTGCTGCTTCGGTATGAAGCTTTTGAAATGACGCGACCATGACTTCT 180  
DB 839 CAGGATATCGGTGCTGCTTCGGTATGAAGCTTTTGAAATGACGCGACCATGACTTCT 898  
QY 181 GGTCAACTGGGAACAGTGGATGTCCACCCATTTGGTGTGCTTCAACAGCAAGA 240  
DB 899 GGTCAACTGGGAACAGTGGATGTCCACCCATTTGGTGTGCTTCAACAGCAAGA 958  
QY 241 TCCTAGTCCCGCCAGGACCATCAATGCAAGTTCACCGATGGAAGGCTTACTCTATGA 300  
DB 959 TCCTAGTCCCGCCAGGACCATCAATGCAAGTTCACCGATGGAAGGCTTACTCTATGA 1018  
QY 301 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGAATTTCCACATCAAGATGGTGGAGACA 360  
DB 1019 AACGGCTGGTGGGTCCAGGACGCTTCCCGTGAATTTCCACATCAAGATGGTGGAGACA 1078  
QY 361 TGAAGTACCCCTTTAGGAGGCGATGGGCTGGAAGTGGTGGACAGTCCAGGTGTAC 420  
DB 1079 TGAAGTACCCCTTTAGGAGGCGATGGGCTGGAAGTGGTGGACAGTCCAGGTGTAC 1138  
QY 421 GCATCTCGATGGCTGTGTGTGACACAGTAATTCGGGGTTCGCTACGCTCTCTACGAGG 480  
DB 1139 GCATCTCGATGGCTGTGTGTGACACAGTAATTCGGGGTTCGCTACGCTCTCTACGAGG 1198  
QY 481 ATGTGTACAGTACAGACATTTCTGTGTCACATGTGTGAGCCCTTATCCACCCAGTGG 540  
DB 1199 ATGTGTACAGTACAGACATTTCTGTGTCACATGTGTGAGCCCTTATCCACCCAGTGG 1258  
QY 541 GTTGTGTACAGTGTGTGGCCACGGCATCAAGATGTCAAGAGCGGAAGTGCATGCGCC 600  
DB 1259 GTTGTGTACAGTGTGTGGCCACGGCATCAAGATGTCAAGAGCGGAAGTGCATGCGCC 1318  
QY 601 ATCAACCCACCTTCGGAAGATCTACTGTGATGCGCTTCTTACTCTTCAAGAGGTAC 660  
DB 1319 ATCAACCCACCTTCGGAAGATCTACTGTGATGCGCTTCTTACTCTTCAAGAGGTAC 1378  
QY 661 GAGCAGTCTACAGAGCGCGTGTGTGTGAGGAAGGATGAAGTGGAGGCGCATTGACC 720  
DB 1379 GAGCAGTCTACAGAGCGCGTGTGTGTGAGGAAGGATGAAGTGGAGGCGCATTGACC 1438  
QY 721 CCCTGAATCTGGGAACATCTCGGTGGCACTGTCTGTAAGGTTCTCTCGATGGATACC 780  
DB 1439 CCCTGAATCTGGGAACATCTCGGTGGCACTGTCTGTAAGGTTCTCTCGATGGATACC 1498  
QY 781 TGATGATCTGTGAGACGGGGGCGCTTCCACAGATGCTTGGAGTGTCTGTACCATG 840  
DB 1499 TGATGATCTGTGAGACGGGGGCGCTTCCACAGATGCTTGGAGTGTCTGTACCATG 1558  
QY 841 CCTCTTCCAGCCCATCTTCCCGGCCACCTTCTGTGAGAGATGATGACTGAGCTACAC 900

1559 CCTCTTCCACGCGCATCTTCCCGGCGCACCTTCTGTGTCAGAGAATGACATTGAGTCACAC 1618  
QY 901 CGCRAAAGGTTATGAGGACAGACTTTCAACTGGGAGAACTACTTGGAGAGAGCAACACT 960  
DB 1619 CGCRAAAGGTTATGAGGACAGACTTTCAACTGGGAGAACTACTTGGAGAGAGCAACACT 1678  
QY 961 CGAAGCCGCTCCATCGAGACTCTTTAAACATGATGTCGCAACCACTGCTTCAAGGTGG 1020  
DB 1679 CGAAGCCGCTCCATCGAGACTCTTTAAACATGATGTCGCAACCACTGCTTCAAGGTGG 1738  
QY 1021 GCATGAAGCTGGAGCGCTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1080  
DB 1739 GCATGAAGCTGGAGCGCTGGACCTGATGGAGCCCGGCTCATCTGTGTGGCCACGGTGA 1798  
QY 1081 AACGAGTGTGATCGGCTCTTCCAGCATCCACTTTTACGGCTGGGACGAGTACGACC 1140  
DB 1799 AACGAGTGTGATCGGCTCTTCCAGCATCCACTTTTACGGCTGGGACGAGTACGACC 1858  
QY 1141 AGTGGTGTGACTGGAGTCCCGACAGATCTACCCGCTGGCTGTGTGAGCTCAGCGCT 1200  
DB 1859 AGTGGTGTGACTGGAGTCCCGACAGATCTACCCGCTGGCTGTGTGAGCTCAGCGCT 1918  
QY 1201 ACCAGTCCAGCTCTCTGTGGCGGACAGAACCGGCCACAC 1239  
DB 1919 ACAGCTCCAGCTCTCTGTGGCGGACAGTGTGGGCTCTC 1957

BC030864 3395 bp mRNA linear ROD 16-APR-2003  
Mus musculus RIKEN cDNA 4732493N06 gene, mRNA (cDNA clone MGC:31247  
IMAGE:4211489), complete cds.  
BC030864  
BC030864.1 GI:21410117  
MGC.  
Mus musculus (house mouse)  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 3395)  
Strausberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,  
Klausner, R.D., Collins, P.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,  
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,  
Hopkins, R.F., Jordan, H., Moore, T., Max, J., Wang, J., Hsieh, F.,  
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Boutford, G.C., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Zzywinski, M.I., Skalska, U., Smal, D.E.,  
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
22388257  
PUBMED  
12477932  
2 (bases 1 to 3395)  
Strausberg, R.  
Direct Submission  
Submitted (03-JUN-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk

RESULT 10  
BC030864  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT



QY 1321 GACCCCTCAGACAGGGTCCAGAGAGCCCTGCTGGAGGACGACCCCTCAGGCTGCCAGGA 1380  
 Db 1931 GGCCCTCAGACAGGGTCCAGAGAGCCCTTACTGGAGGACAACTTGAGGCTTTGGGG- 1989  
 QY 1381 AGATCTCGTGGAGCCTGCTCCCTGGCGAGATCATTTGCTGCGTGTGAGGAGAGACATC 1440  
 Db 1990 -----GTCCTGGAACCACTGCTTCCTGATGACATTTATCGCTGTGTGTGAGGAGGAGGCACC 2044  
 QY 1441 TAGACGTGGCTCCGCCGACAGAGCTTCAAGTCCAGAGCTGCTGCTGCTCCGTCGAGAAC 1500  
 Db 2045 AGGACATTTCTCGCTGACAGAGTCCGCCAGTCCACAGCTGCTCTCCCATTTGAGAGCA 2104  
 QY 1501 TCAGCAGGAGAAACAGACATGAGCCTT-CTGCTCCAGCCTGGCTTCTAGCTGGAAGC 1559  
 Db 2105 TCAGCAGGAGAGAGAACTGAGACTTCCCTGGCATCAGCCTGGACCTTAAGTGAAGCC 2164  
 QY 1560 CAGC 1563  
 Db 2165 AAGC 2168

RESULT 11  
 HS756G23  
 LOCUS  
 DEFINITION  
 Human DNA sequence from clone RP4-756G23 on chromosome 22q13.31-13.33 Contains the 5' part of a gene similar to drosophila transcriptional repressor, the 3' end of the gene for a novel Leucine Rich Protein, the RANGAPI gene for Ran GTPase activating protein 1, ESTs, STS, GSSs and three putative CpG islands, complete sequence.  
 ACCESSION  
 AL035681  
 VERSION  
 AL035681.13 GI:4902689  
 KEYWORDS  
 HNG; CpG island; Ran GTPase activating protein; RANGAPI.  
 SOURCE  
 Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 89948)  
 Direct Submission  
 Submitted (11-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk  
 On May 27, 1999 this sequence version replaced gi:4775627.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WormPeP; Information on the WormPeP database can be found at <http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the 'unseq' feature key. This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>  
 RP4-756G23 is from the library RPCI-4 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pCYPAC2  
 IMPORTANT: This sequence is not the entire insert of clone RP4-756G23 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-756G23 is at 1 in this sequence. The

true left end of clone RP5-979N1 is at 89849 in this sequence. The true right end of clone RP1-85F18 is at 80441 in this sequence.

FEATURES  
 Location/Qualifiers  
 1. 89948  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="22"  
 /map="ql3.31-13.33"  
 /clone="RP4-756G23"  
 /clone\_lib="RPCI-4"  
 repeat\_region  
 2172..2465  
 /note="Alusg1 repeat: matches 1..296 of consensus"  
 complement(10832..11352)  
 /note="match: GSS: Em:AQ266638"  
 misc\_feature  
 complement(10989..11290)  
 /note="match: GSS: Em:AQ269066"  
 misc\_feature  
 11441..11790  
 /note="match: GSS: Em:AQ204186"  
 misc\_feature  
 11455..11594  
 /note="match: GSS: Em:AQ285637"  
 17446..24698  
 /gene="dJ756G23.3"  
 join(17446..17677,20088..20170,20515..20749,21230..21295,22956..23082,23565..24698)  
 /gene="dJ756G23.3 (novel protein similar to drosophila transcriptional repressor)"  
 /note="match: ESTs: Em:H22704 Em:AA351643 Em:AI364341 Em:T995339 Em:R06448 Em:R36553 Em:R17171 Em:D76919 Em:AA066436 Em:A1459129 Em:A1660728 Em:AA719364 Em:AI361918 Em:A1587146 Em:AI150426 Em:AI160979 Em:AI522858 Em:A1537859 Em:AA281766 Em:AA233695 Em:R27062 Em:AA905213 Em:R32930 Em:AA356192 Em:AA026761 Em:AA353188 Em:AA280121 Em:NS6014 Em:TS8551"  
 /evidence=not experimental  
 CDS  
 join(17446..17677,20088..20170,20515..20749,21230..21295,22956..23082,23565..23677)  
 /gene="dJ756G23.3"  
 /note="match: proteins: Tr:O75996 Tr:Q24191 Wp:CE06250 Tr:O76931"  
 /codon\_start=1  
 /evidence=not experimental  
 /product="dJ756G23.3 (novel protein similar to drosophila transcriptional repressor)"  
 /protein\_id="CAB63071.1"  
 /db\_xref="GI:6572271"  
 /db\_xref="SPTREMBL:Q9UGS4"  
 /translation="MVESMKYPRQGMRLVVDKSVSRTRMAVVDTVIGRLRLLYE DQSDDDDFWCHMSPLIHPVGSRRVGHGKMSGYEAQTENWENYLEKSKAAPSRL FNMDPNHGFVKMGKLEAVDLMEPRLICVATKRVVHRLLSIHFDGWDSEYDQWDC SPDIYPVGCETGYQLQPPVAEPATPLAKATKKKKQFGKKRIIPTKTRPLR QGSKKPLEDDPQARKISSEPVPEGIIVRVKHEHLDVASPKASPELPVSVENIK QETDD"  
 complement(24311..24697)  
 /note="match: STS: Em:G14475"  
 misc\_feature  
 complement(24350..24620)  
 /note="match: STS: Em:G43605"  
 misc\_feature  
 complement(24449..24697)  
 /note="match: STS: Em:G14931"  
 24698  
 /gene="dJ756G23.3"  
 complement(join(28566..28769,29910..30076,30602..32311,32869..33046,34152..34282))  
 /gene="dJ756G23.1"  
 complement(join(28566..28769,29910..30076,30602..32311,32869..33046,34152..34282))  
 /note="match: CDNAS: Em:U08018 match: ESTs: Em:AA128077 Em:AA326017 Em:W58996 Em:AI460232 Em:AA367412 Em:AA126901 match: proteins: Tr:O55226 Sw:O02833 Sw:P22792 Sw:Q01129 Tr:O86637"

```
/codon_start=1
/evidence=not_experimental
/product="dj756G23.1 (novel Leucine Rich Protein)"
/protein_id="CAB63072.1"
/db_xref="GI:6572272"
/db_xref="SPTREMBL:Q9UGS3"
/translation="MVQGGIQSSVASGKSLPPLGLSLBAGGQGLMGLPVLQEGGLPR
STSHVPLVLLVLLAPAAQACQPOAC:CDNSRRHVACRYQNLTEVDAIP
EUTQRDLQGNLKVTPAAAFQGVFHLTHLDHCEVBLVAEAFGLGRLLNLNLAS
NHLRELPOALDGLSGLEGNALBELRPTGTCALGALATNLNAHNLVYLPA
FOGLRVRLRSHNALSVALEALAGLPALRRLSLHNEQLALPVLQAGRLARL
ELGNPLTVAGSEEDGLALPGLRELLLDGALCALGPRAPFACPLRHLTLDLGNQDLT
PFLQGGQRRLRVCWCCQARPLLEWLRARVSDGACQFDRLRGRGADALR
PMDLRCPGAAGAEELERAVAGRPFPGRGGERAVAPCPACVCPVPSRHS
SCGGCLQVAPGSDTQLDLRRNRPSPVPAFPGHLVLSLHLCQGLAELEG
ALAGRLLYLISNQLAGSAALEGAAPRGLYGLERNFLQVFGAALRALPSLFS
LHLDNVRAPDGLRTRALRWVYLSGNRTVEVSLGALGPARELKLHLDNRQRE
VPTGALGLPALLELOLSNPLRALRDGAFQVGHSLQHLFLNSSLGLRQICPGAFSG
FGQLQSLHKKQLRALPALPSLSQLELIDLSNPPHCDQLLPLRWLTGLNLRYGA
TCAIPNPAQQRVAAAFAVEDPCGWAARKRTFASRPTFIRKQCGADKVG"
/notes="CpG island"
/evidence=not_experimental
/complement(32183..32620)
/gene="dj756G23.1"
/notes="match: GSS: Em:AQ389159"
32627..33079
/notes="match: GSS: Em:AQ383531"
35161..35287
/notes="AluJo/FLAM repeat: matches 1..131 of consensus"
35563..35664
/notes="51 copies 2 mer ta 86% conserved"
36802..36970
/notes="MER9A repeat: matches 1..178 of consensus"
complement(39025..79639)
/gene="RANGAP1"
complement(join<39025..40098,42756..42877,43155..43243,
44433..44535,46298..46417,47734..47920,49447..49531,
49617..49716,50137..50250,51374..51532,54872..55006,
58090..58269,61523..61582,68026..68153,74359..74555,
79412..79639))
/gene="RANGAP1"
/product="dj756G23.2 (Ran GTPase activating protein 1)"
/notes="match: cDNAs: Em:U08110 Em:X82260 Em:U20857
Em:U88155 Em:AB002306 Em:D79984 Em:X07648
match: ESTs: Em:AA435922 Em:AA280707 Em:AA351030
Em:AA789231 Em:AA985159 Em:H46148 Em:T08060 Em:H21130
Em:T08061 Em:AA026670 Em:AA412337 Em:T35912 Em:D82175
Em:A1139769 Em:A1203851 Em:AA852745 Em:AA412447
Em:AA026631 Em:R26090 Em:H17609 Em:H47302 Em:AA991855
Em:T30135 Em:N90379 Em:AA769102 Em:A1138957 Em:AA852746
Em:AA52593 Em:H18108 Em:H18150 Em:W52172 Em:AA324880
Em:R78185 Em:H68027 Em:AA343282 Em:AA383824 Em:H40643
Em:H50021 Em:AA371797 Em:H55068 Em:D25677"
/evidence=not_experimental
/complement(39025)
/gene="RANGAP1"
39047..39317
/notes="match: STS: Em:G20672"
complement(39054..39059)
/gene="RANGAP1"
complement(39301..39445)
/notes="match: STS: Em:G44270"
complement(join(40029..40098,42756..42877,43155..43243,
44433..44535,46298..46417,47734..47920,49447..49531,
49617..49716,50137..50250,51374..51532,54872..55006,
58090..58269,61523..61582,68026..68153,74359..74470))
/gene="RANGAP1"
/notes="match: proteins: Sw:P46060 Sw:P46061 Sw:O13066
Tr:O17437 Tr:O36421"
/codon_start=1
/evidence=not_experimental
```

```
Query Match
Best Local Similarity 99.9%; Score 1132.4; DB 9; Length 89948;
Matches 1133; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1411 TCATTGCTGTGCGTGTGAAGGAAGAGCATCTAGACGTGGCTCGCCGCGCAAGAGCTTCAA 1470
DB 23565 TCATTGCTGTGCGTGTGAAGGAAGAGCATCTAGACGTGGCTCGCCGCGCAAGAGCTTCAA 23624
QY 1471 GTCAGAGCTGCCTGTCTCCGTGAGAACATCAAGCAGGAGAAACAGACGACTGAGCCTTCC 1530
DB 23625 GTCAGAGCTGCCTGTCTCCGTGAGAACATCAAGCAGGAGAAACAGACGACTGAGCCTTCC 23684
QY 1531 TGCTCCAGCTGGCTTCTAGCTGGAGCCAGCCAGCGCTTCTTCTACCAACACACCAT 1590
DB 23685 TGCTCCAGCTGGCTTCTAGCTGGAGCCAGCCAGCGCTTCTTCTACCAACACACCAT 23744
QY 1591 GCCTCCACCTGACTTTGGCTTGGAGACTGATCTCTCTGTGTAAATTTCTGCCCGTGTCTG 1650
DB 23745 GCCTCCACCTGACTTTGGCTTGGAGACTGATCTCTCTGTGTAAATTTCTGCCCGTGTCTG 23804
QY 1651 TGAAGGCTGACCGGTGGAGGACCTGCTGGGGTCTCTGGGACCGCGCTGTGTCTTCTGCC 1710
DB 23805 TGAAGGCTGACCGGTGGAGGACCTGCTGGGGTCTCTGGGACCGCGCTGTGTCTTCTGCC 23864
QY 1711 CTCCTCTGTGGAAGGTCTATATGACGGGCGGCTGAGGCCCGCCAGAACTCGTCTGTGAAC 1770
DB 23865 CTCCTCTGTGGAAGGTCTATATGACGGGCGGCTGAGGCCCGCCAGAACTCGTCTGTGAAC 23924
QY 1771 CACTTTTTCAGGCAGAGTTCACAAAGCTGGAAAGCTAGTGCCTGCTTCTTCTTAAAGAT 1830
DB 23925 CACTTTTTCAGGCAGAGTTCACAAAGCTGGAAAGCTAGTGCCTGCTTCTTCTTAAAGAT 23984
QY 1831 GGCTCTCCCGGACCGCCACGCGCTCAGTTCGCCAGGAGTGGGCCACCACTGTGCACAC 1890
DB 23985 GGCTCTCCCGGACCGCCACGCGCTCAGTTCGCCAGGAGTGGGCCACCACTGTGCACAC 24044
QY 1891 TGTGGAAATACAAGACAGTGAACCTGTCTGCCCTGAACAGACTCATGTAAATTAAGTTCTAG 1950
DB 24045 TGTGGAAATACAAGACAGTGAACCTGTCTGCCCTGAACAGACTCATGTAAATTAAGTTCTAG 24104
QY 1951 AGCAGCTCTCTGACAGGATTAAGTCTCCCTGACAGTGAAGTGTGTGGTGGGGGACGCTC 2010
DB 24105 AGCAGCTCTCTGACAGGATTAAGTCTCCCTGACAGTGAAGTGTGTGGTGGGGGACGCTC 24164
QY 2011 TGCCTCAAAAATTCACCAAGCAGAAATGCTCTCTCAGCCTCATGTGTGGTCTCTCTGCTCCT 2070
DB 24165 TGCCTCAAAAATTCACCAAGCAGAAATGCTCTCTCAGCCTCATGTGTGGTCTCTCTGCTCCT 24224
QY 2071 CCTAGCTCCCGAGGATGTTGGGGACCCAGCTGTCTCGCAGCTAAGAGAGAGTGAACCA 2130
DB 24225 CCTAGCTCCCGAGGATGTTGGGGACCCAGCTGTCTCGCAGCTAAGAGAGAGTGAACCA 24284
QY 2131 GGATGTGATTTTGGCGACCTGTGTGGTGGCTTTGAGCTGCTTCTGTGTGTGTGTGAGGAC 2190
DB 24285 GGATGTGATTTTGGCGACCTGTGTGGTGGCTTTGAGCTGCTTCTGTGTGTGTGAGGAC 24344
QY 2191 TGACTCCCATTTCTAAAGAAATGCCCCGGGGAGGACATTTGGAGAGGAAGATGGCCCTGA 2250
DB 24345 TGACTCCCATTTCTAAAGAAATGCCCCGGGGAGGACATTTGGAGAGGAAGATGGCCCTGA 24404
QY 2251 GTGTGCACTTTGGCTCTGCTACCTGCTCCCTGAGACCCCGCTTAAATAATTAATTCATCCAAGA 2310
DB 24405 GTGTGCACTTTGGCTCTGCTACCTGCTCCCTGAGACCCCGCTTAAATAATTAATTCATCCAAGA 24464
QY 2311 TTCCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAGCTGGGCTTGTATG 2370
DB 24465 TTCCTTTGTAGTTAAAGGGTCCAGTTCTGACTGGAGCCTCTAGAGAGCTGGGCTTGTATG 24524
```



QY 2371 TTCTTTGGCCCTTTGTTTCTTACCTAAATGAAGAAACCATGCTGGAGGGCCCTGGAACA 2430  
 Db 24525 TTCTTTGGCCCTTTGTTTCTTACCTAAATGAAGAAACCATGCTGGAGGGCCCTGGAACA 24584  
 QY 2431 CAGAACCTTCAGACCAAGGATGACAGCTGGAGGACACATCTAGCTGCAATTCGAACCT 2490  
 Db 24585 CAGAACCTTCAGACCAAGGATGACAGCTGGAGGACACATCTAGCTGCAATTCGAACCT 24644  
 QY 2491 CACTGGGCTCCCGACAGCTCTGTGTGTGAGAAATTAACCCCTGCTTCTTTGAG 2544  
 Db 24645 CACTGGGCTCCCGACAGCTCTGTGTGTGAGAAATTAACCCCTGCTTCTTTGAG 24698

RESULT 12  
 AY237001  
 LOCUS  
 DEFINITION Mus musculus M4MBT variant B (M4mbt) mRNA, complete cds;  
 alternatively spliced.  
 ACCESSION AY237001  
 VERSION AY237001.1 GI:29569825  
 KEYWORDS  
 SOURCE Mus musculus (house mouse)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 REFERENCE 1 (bases 1 to 3399)  
 AUTHORS Markus,J., Feikova,S., Sramko,M., Wolff,L. and Bies,J.  
 TITLE Cloning and molecular characterization of the novel murine gene  
 M4mbt encoding a nuclear zinc finger protein with four mbt domains  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3399)  
 AUTHORS Markus,J., Feikova,S., Sramko,M., Wolff,L. and Bies,J.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-FEB-2003) Department of Molecular Virology, Cancer  
 Research Institute of Slovak Academy of Sciences, Vlnarska 7,  
 Bratislava, Slovak Republic  
 FEATURES  
 source  
 1..3399  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="Swiss Webster"  
 /db\_xref="taxon:10090"  
 gene  
 1..3399  
 /gene="M4mbt"  
 CDS  
 46..2187  
 /gene="M4mbt"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="M4MBT variant B"  
 /protein\_id="AA084917.1"  
 /db\_xref="GI:29569826"  
 /translation="MEKPRGTETAPSPSEPEEEDLDLFGGYDPSRYSNSAGSES  
 SSYLESEANEDEAREGELPTSEHLFSSANNRSLDGSSEPAVCEMCGIVGTREAP  
 FSKTRFCVSRSYSNSKASILARLOQPKTKAKVLHKAASAKIGAFILHAGG  
 TQADAGPTQDALVDFDRGKFLKHSYKAAPVSCFKHPLYDOWEDYMGKMKVEV  
 LNSADVLPSTVYVATVIAQAGYVLLRYEGFENDSHDFWNLGTVDVHPHIGWCAIN  
 SKILVPPRTIAKT'DWKSILMKLNGSRTLPA'DPHIKVYESMKYPRQGRLEVVDK  
 TOVSTRNAVVDTWIGGELRLLYEDGSDDDFWCHMSPLIHPVMSRVCHGILKMSD  
 RCDNSHPTREKLYCDAPVPLFKRYAVYTEGWFEGMKLEAIDPLNLGSCVATI  
 CKVLDDGYLIMICDVGSPSTGSDWFCYHASSHAIFPATFOKNDIELTTPRGYETQPF  
 AWETYLEKTSKAPALFNWDCNDPCPHGFKVGMKLEAIDPLNLGSCVATI  
 SIHFDGWNEDYDWDCESDPIYPVWCELTGYQLQPPVSAEPNTPQKXDTTKKKKK  
 QFGKKRRIPSAKRIPRLRQSGSKPLLEDNLALGVSPVPDPSRPVHSCLPPLRASSR  
 RGTETSLASAWTLTEAKHGAERREGFTAVQ"  
 313..357  
 /gene="M4mbt"  
 /note="Region: putative zinc finger"  
 misc\_feature  
 580..894  
 /gene="M4mbt"  
 /note="Region: mbt repeat"  
 misc\_feature  
 916..1212  
 /gene="M4mbt"

misc\_feature  
 1267..1545  
 /gene="M4mbt"  
 /note="Region: mbt repeat"  
 misc\_feature  
 1567..1857  
 /gene="M4mbt"  
 /note="Region: mbt repeat"  
 BASE COUNT 825 a 905 c 947 g 722 t  
 ORIGIN  
 Query Match 44.2%; Score 1128.2; DB 10; Length 3399;  
 Best Local Similarity 88.9%; Pred. No. 3.9e-278; Indels 0; Gaps 0;  
 Matches 1220; Conservative 0; Mismatches 153;  
 QY 1 AGCACATCCCACTCTATGACCAAGTGGAGGATGTGATGAAAGGATGAAGGTTGAGGTGC 60  
 Db 641 AACATGTACCCCTCTATGACCAAGTGGAGGATGTGATGAAAGGATGAAGGTTGAGGTGC 700  
 QY 61 TCACACAGTGTGCTGTGCTCCCGAGCGGGTGTACTGATCGCCTCTGTATCCACAGAG 120  
 Db 701 TCAACAGCGATGCTGTGCTCCCGAGCGGGTGTACTGATCGCCTCTGTATCCAGGCAG 760  
 QY 121 CAGGGTATCGGGTGTGCTTCCGATGAAAGGCTTTGAAATGACGCCAGCCATGACTTCT 180  
 Db 761 CTGGGTACCGGGTGTGCTTCCGATGAAAGGCTTTGAAATGACGCCAGCTCATGACTTCT 820  
 QY 181 GGTGCAACCTGGGAAACAGTGGATGTCCACCCATTGGCTGGTGTGTCATCAACAGCAAGA 240  
 Db 821 GGTGCAACCTGGGAACTGTGGATGTCCACCCATTGGTGTGTCATCAACAGCAAGA 880  
 QY 241 TCCTAGTGTGCTGCTGCTCCCGAGCGGCATCCATGCAAGTTCAACGACTGGAAGGGCTCCTCATGA 300  
 Db 881 TCCTGTGACTTCCACGAGACATCCATGCAAGTTTACTGACTGGAAGAGACTACCTCATGA 940  
 QY 301 AACGGTGTGGGCTCCAGAGAGCTTCCGTTGGATTTCCATCAAGATGGTGGAGAGCA 360  
 Db 941 AGCGGTGTGGTGTGCTTCCAGAGAGCACTTCTCCGAGACTTCCATATCAAGATGGTGGAAAGCA 1000  
 QY 361 TGAAGTACCCCTTTAGGCAAGGCGATCGCGTGGAGTGTGGACAGTCCCGAGGTGTCTAC 420  
 Db 1001 TGAAGTACCCCTTTCCAGAGGCGATCGCGTTAGAGTTGTAGCAAGATCTAGGTGTCTAC 1060  
 QY 421 GCACCTCGCATGCTGTGTGGACACAGTAATCGGGGGTCCGCTTACGGCTCCTCTACGAGG 480  
 Db 1061 GTACCCGCAATGCGCGTGTGGACACAGTAATCGGGGGTCCGCTCCGGCTCCTCTATGAGG 1120  
 QY 481 ATGGTGACAGTACGACGACTTCTGTGTGCCACATGTGGAGCCCCCTGTATCCACCCAGTGG 540  
 Db 1121 ATGGTGACAGTGTATGACGACTTCTGTGTGCCATATGTGGAGTCCCTGTATCCACCCAGTGG 1180  
 QY 541 GTTGGTCAAGAGTGTGGGCCACGGCATCAAGATGTCAGAGAGGCGAAGTGCATGGCCCC 600  
 Db 1181 GTTGGTCCGGGGTGTGTGGCCACGGCATCAAGATGTCAGACAGATGTGACATGTCTC 1240  
 QY 601 ATCACCCCACTTCCGGAAAGATCTACTGTGATGCGCTTCCCTTACCTTTTCAAGAGAGTAC 660  
 Db 1241 ATCACCCCACTTCCGGAAAACTACTGTGATGCTGTACCTTTACCTTCTTCAAGAGAGTCC 1300  
 QY 661 GAGCAGTCTACACAGAGCGGGTGTGTTTGGAGAGGAGTGAAGTGGAGGCCATTGACC 720  
 Db 1301 GCCTGTCTACACAGAGGTGGTGTGTTCCGAGAGAGGAGTGAAGTGGAGGCCATTGACC 1360  
 QY 721 CCCTGAATCTGGGCAACATCTCGGTGGCAACTGTCTGTAAAGTGTCTCTGTGATGGATACC 780  
 Db 1361 CTCTGAATCTGGGCACTATCTGTGTAGCAACCATCTGCAAGGTGCTCTTGGATGGTTACC 1420  
 QY 781 TGAATGATCTGTGGACGGGGGGCCCTCCACAGATGGCTTGGAGTGTCTCTGCTACCATG 840  
 Db 1421 TGAATGATCTGTGGATGGGGGGCCCTCCACGAGTGGCTCCGATTTGTTCTCTACCCAGC 1480  
 QY 841 CCTCTTCCACGCCATCTTCCCGGGCCACCTTCTGTGAGAGAGTATGACATTTGAGCTCACAC 900  
 Db 1481 CCTCTTCCCATGCCATCTTCCCGACCCACCTTCTGCGCAAGAGTATGACATTTGAGCTCACAC 1540





QY 2292 AAAAATAATTCATCCAGATTCCTTTGTAGTAAAGGTCAGTCTTGACTGGAGCTCT 2351  
 Db 361 AAAAATAATTCATCCAGATTCCTTTGTAGTAAAGGTCAGTCTTGACTGGAGCTCT 420  
 QY 2352 AGAGAGCTGGGCTGTATGTTCTTTGGCCCTTTTGTTC 2389  
 Db 421 AGAGAGCTGGGCTGTATGTTCTTTGGCCCTTTTGTTC 458

RESULT 15  
 LOCUS AX046388 3590 bp DNA linear PAT 24-NOV-2000  
 DEFINITION Sequence 55 from Patent WO0011168.  
 ACCESSION AX046388  
 VERSION AX046388.1 GI:11344369  
 KEYWORDS Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1  
 Lemischka, I. and Moore, K.  
 Genes that regulate hematopoietic blood forming stem cells and uses  
 thereof  
 JOURNAL Patent: WO 0011168-A 55 02-MAR-2000;  
 Princeton University (US)  
 FEATURES  
 source Location/Qualifiers  
 1..3590  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"

BASE COUNT 1101 a 702 c 797 g 990 t

ORIGIN  
 Query Match 16.2%; Score 413.8; DB 6; Length 3590;  
 Best Local Similarity 60.2%; Pred. No. 8.5e-95;  
 Matches 735; Conservative 0; Mismatches 462; Indels 24; Gaps 2;

QY 1 AGCAGTCCCACTCTATGACAGGAGGAGTGTGTAAGGGATCAAGGTGAGGTGC 60  
 Db 750 AGCATGACCTATGGGACCTGCTGGGGTATATCTCGGAAATGTAAGGATAGAGTTC 809  
 QY 61 TCAACAGTGTGCTGTCTCCAGCCGGGTGTAAGGATCGGCTCTGTATCATCAGACAG 120  
 Db 810 CCAATACAGAGCTGAGTCTACCTACCAAGTCTCTGATGCTGGAATTAATAATTAG 869  
 QY 121 CAGGTATCGGGTCTGCTGCTGATGAGGCTTTGAAATGAGGCGGACGATCTCT 180  
 Db 870 CAGGTATTAATGCTCTTTTGAGATATGAGGATTTGAAATGATTTCTCTGAGCTTCT 929  
 QY 181 GGTCAACCTGGGAACAGTGTATCCACCCCATTTGGCTGTGTGCCATCAACAGCAAGA 240  
 Db 930 GGTGCAATATATGTTGGGTCTGATATTTATCATCAGTGTGTTGGTGTGAGTAGTGGAAAC 989  
 QY 241 TCTAGTGTCCCAAGGACCATTCATGCCAAGTTTACCGATCGGAGGGTCTCTCATGA 300  
 Db 990 CTCTCGTTCCTCTAGAACTGTTCAACATAAATATACAACTGGAAGCTTTCTAGTAA 1049  
 QY 301 AACGGTGTGGCTCCAGGAGGCTTCCCTGGATTTCCACATCAGATGGTGGAGCA 360  
 Db 1050 AAAGACTTACTGTGGTCCAAACACTTCTCTGATTTTCTCAAAAAGTTTCTGAGAGTA 1109  
 QY 361 TGAAGTACCCCTTTAGGAGGAGGATGGGCTGGAAGTGGTGGACAGTCCAGGTGTCTAC 420  
 Db 1110 TGCAATATCTTTCAAACTTGTGATGAGTGTAGAGTAGTTGACAGAGGCATTTATGTC 1169  
 QY 421 GCATCGCATGCTGTGTGTGACACAGTATTCGGGGTTCGCTACGGCTCTCTACGAGG 480  
 Db 1170 GAAACAAGATGAGTGTGTGAAAGTGTAAATTGAGGAGCAGTACGGCTGTGTATGAAG 1229  
 QY 481 ATGGTGCAGATG---ACGACGACTCTGTGGTCCACATGTGGAGCCCTCTGATCCACCCAG 537  
 Db 1230 AGAGTGAAGATGGACAGACGACTTCTGTGGTCCACATGACAGGCCCTTAAATCCACATA 1289

Search completed: February 4, 2004, 16:20:37  
 Job time : 9222 secs

QY 538 TGGTGTGTCAGCGTGTGGGCCACGGCATCAAGATGTCAAGAGGCGAAGTGCATGG 597  
 Db 1290 TTGGATGTCAAGAAGCATAGCCCATCGATTCAAGAGATCAGATATTACGAAGAAACAGG 1349  
 QY 598 CCATCACCACCCACCTTCGGAAGATCTACTGTGATGCGGTTCCTTACCTCTTCAAGAGG 657  
 Db 1350 ACGGAC-----ATTTTCGATACACCTCCACACTTATTTCGTAAGG 1388  
 QY 658 TACGAGCAGTCTACACAGAAGCGCGTGTGTTGAGGAGGATGAAGCTGGAGGCCATTG 717  
 Db 1389 TAAAGAGTAGACACAGAGTGGAGATGGTTCAAGAGAGGAATGAAATTGGAAGCTATAG 1448  
 QY 718 ACCCTCTGAATCTGGGCAACATCTGCTGGGCACTGTCTGTAAAGTCTCTCTGTGATGAT 777  
 Db 1449 ACCCATTAATCTTTCTTACAATATGTGTGCCACCATTAGAAAGGTGTGGCTCATGGAT 1508  
 QY 778 ACTCATGATCTGTGTGACGCGGGGCCCTCCACAGATGGCTTGAGTCTGTGCTTCTGCTACC 837  
 Db 1509 TCCTGATGTTGGGATTTGATGGCTCAGNAGCAGCAGATGGATCTGACTGGTCTGTATC 1568  
 QY 838 ATGCTCTTCCACGCGCATCTTCCGCGCCACTTCTGTCAAGAAATGACATTTGAGTCA 897  
 Db 1569 ATGCAACCTCTCTCTCCATTTTCCCTGTGGGTCTGTGAAATTAACATGATAGAACTGA 1628  
 QY 898 CACGCGCAAAAGGTTATGAGGCAAGACTTTCAACTGGGAACTACTTTGGAGAGACCA 957  
 Db 1629 CTCACCCAGAGGTTACACAAACTTTCCTTTTAATGTTGACTTACCTCAGGGAACCG 1688  
 QY 958 AGTCGAAAGCGCTCCATCGAGACTCTTTAAACATGGATTGCCAAACCATGGCTTCAAGG 1017  
 Db 1689 GCTCCATTGACAGCACCAGTAAACTATTTAATAAGGATGTTCCAAACCCAGGATTCGGT 1748  
 QY 1018 TGGGATGAAGCTGGAGGCCGTGACCTGTAGTGGAGCCCGGCTCATCTGTGTGSCCAGG 1077  
 Db 1749 TAGGAATGAAATTAAGAGCTGTAGATCTCATGGAGCCACGGTTAATATGTGTAGCCACAG 1808  
 QY 1078 TGAACAGAGTGTGTCATCGGCTCTCAGCATCCACTTTGACGGCTGGGACAGCAGTACG 1137  
 Db 1809 TTACTCGAATTTTACCCTCTCTGAGGATACATTTTGTGTTGGGAGAGAGTATG 1868  
 QY 1138 ACCAGTGGGTGAGTGGAGTCCCGCAGACATCTACCCCGTGGGTGTGTGAGTGTCAAG 1197  
 Db 1869 ACCAGTGGGTGAGTGTGAGTCCCGTGTGAGTCCCTGACCTCTATCTGTAGGGTGTGTCACTG 1928  
 QY 1198 GCTACCCAGCTCCAGCCTCCTG 1218  
 Db 1929 GATATCAACTACAGCCTCCAG 1949